The Development of Morality

Robert James Richard Blair

Department of Psychology,
University College,
London.

Thesis submitted for
the degree of Ph.D. in Psychology,
University of London.
November, 1992.

Acknowledgements

My greatest thanks must go to John Morton who has supervised this thesis with great patience and enthusiasm. His skilful guidance and rigorous critical discussion have enhanced the quality of this work beyond measure. I am also indebted to Henry Plotkin whose third year seminars were responsible for me attempting this thesis in the first place. In addition, I must thank Paul Burgess, Uta Frith, Annette Karmiloff-Smith, Francesca Happé, Alan Leslie, George Mandler, Jean Mandler, Stuart Miller and Chris Mitchell who have helped me develop some of my ideas. Finally, I thank Jack van Horn and his computing and statistical skills who proved to be an excellent office-mate and Kate Amos and John Draper who rescued me from many unpleasant circumstances.

I am also extremely grateful to Dr. Tidmarsh, Dr. Blackburn, Carol Sellars, Fiona Clarke, Dr. Williams and Dr. Strickland and the nurses and other staff at Broadmoor and Ashworth Special Hospitals. Without their help the studies would no have been possible.

Finally, I thank my parents and Lisa Cipolotti who have been and remain very tolerant.

Abstract

Evidence that individuals distinguish between moral and conventional rules is reviewed. Moral rules prohibit actions that result in victims (e.g., violence, stealing, etc.). Conventional rules prohibit actions that do not result in victims (e.g., not saying please, dressing in opposite sex clothes). Previous theoretical accounts of the development of the moral/ conventional distinction are discussed. These theories are contrasted with an approach that is developed here. It is proposed that there is a mechanism, a Violence Inhibition Mechanism (VIM), that is responsible for the previously observed aversive arousal response to the distress of others. It is proposed that this aversive arousal response is a prerequisite for the development of the moral/ conventional distinction, the moral emotions and the inhibition of violent behaviour.

Previous accounts have stressed role taking as a prerequisite for the moral/ conventional distinction. However, this was found not to be the case. Autistics, already known to be lacking a "Theory of Mind" and therefore unable to role take, were found to make the moral/ conventional distinction.

It was hypothesized that Antisocial Personality Disorder (APD) might be a consequence of a lack of VIM; the clinical description of APD stresses their lack of the moral emotions and their inability to inhibit their violent actions. In line with this, APD subjects were not found to make a moral/ conventional distinction. Two rival explanations of APD were investigated: that APD is due to an inability to role take and; that APD is due to frontal lobe damage. APD subjects were not found to be impaired in either of these respects in comparison to criminal controls.

A final investigation focused on the emotion attributions of APD subjects. It was hypothesized, given the contention that VIM is a prerequisite for the development of the moral emotions, that APD subjects might make anomalous attributions in victim situations though their attributions of other situations should prove normal. This study observed that while the attributions of APD subjects and criminal controls did not differ if the emotions attributed were happiness, sadness or embarrassment there was significant difference in victim situations where APD subjects were less likely to attribute guilt and more likely to attribute indifference than criminal controls. This finding was taken as indirect support of the VIM position.

Additional tests, and implications of the VIM model are then discussed.

Table of Contents

Chapter One The Moral/ Conventional Distinction

Chapter Two A Model of the Development of Morality

Chapter Three Autism and Morality

Chapter Four Antisocial Personality Disorder and Morality

Chapter Seven Antisocial Personality Disorder: A lack of VIM?

Chapter Eight Conclusions and Future Directions

List of Tables

List of Figures

Chapter One

The Moral/ Conventional Distinction

1.1: Introduction

Until now, the most important observation concerning the development of morality has been that children judge moral and conventional transgressions differently. This difference in subject's judgements of moral and conventional transgressions will be termed here *the moral/ conventional distinction.* I will describe the form of this distinction in section 1.2, detail its development in section 1.3 and discuss its critics in section 1.4. In sections 1.5 to 1.7, I will discuss the implications of the distinction for the various theories of the development of morality. However, before I do any of this, I will give the definitions of moral and conventional rules that have been used previously in the research.

According to Turiel (1983a), a set of criteria for the moral domain may be inferred from the philosopher, Gewirth's (1978) orientation. Gewirth states that moral prescriptions are obligatory, universally applicable (in that they apply to everyone in similar circumstances), impersonal (in that they are not based on individual preferences or personal inclinations) and are not determined by consensus. Moral prescriptions stem from concepts of welfare, justice and rights (Dworkin, 1978; Rawls, 1971).

Morality is instantiated by actions (e.g. hitting and hurting) that have an objective effect on the well-being and rights of others. Actions or events possess a moral quality if they involve physical or psychological harm, personal or private property, promises or commitments, or the allocation of scarce resources (Turiel, Killen & Helwig, 1987). Moral prescriptions are not perceived to be alterable by consensus. They are only alterable by changes in their perceived moral merits. Turiel (1983b; p. 39) states:

> "In contrast with conventions, moral prescriptions are not perceived to be alterable by consensus. That is not to say that morality is fixed and unalterable. We know, for instance, that historical changes have occurred with regard to such matters as slavery. However, the bases for these changes are not perceived as shifts in the general consensus or in social organization, but on the intrinsic merits, from the moral point of view, of one type of action over another. Again, in contrast with convention, in the moral domain actions are not arbitrary, and though moral prescriptions form part of social organization, they are not defined by social organization nor is their rationale based on their status as implicit or explicit regulations. The individual's moral

> prescriptions (e.g., regarding killing and the value of life) are determined by factors inherent to social relationships, as opposed to a particular form of social organization. An individual's perception of an act such as the taking of life as a transgression is not contingent on the presence of a rule, but rather stems from factors intrinsic to the event (e.g., from the perception of the consequences to the victim). This means that moral issues are not perceived as relative to the societal context. The moral theories formed by individual's are based on concepts regarding the welfare of persons, the rights of persons, and justice, in the sense of comparative treatment of individuals and means of distribution."

In contrast, Turiel (1983b) states that social contentions are:

> "behavioral uniformities which coordinate interactions of individuals within social systems. Individual members of society have shared knowledge about conventions.... Consequently, conventions (e.g., modes of greeting, forms of address) provide people with means of knowing what to expect of each other and thereby serving to coordinate interactions between people.... Conventions involve coordinations at the level of social organization; they are uniformities that coordinate the stable interactions of individuals functioning within a social system and the ends are social organizational.
>
> Social conventional acts are symbolic elements of social organization. As such, the acts, in themselves, are arbitrary and alternative courses of action can serve similar functions. That is, by virtue of their shared knowledge, a given conventional uniformity in one social system may serve the same symbolic function as a different uniformity in another social system.... Conventions are validated by consensus and, therefore, are relative to the societal context. In addition to the variability of conventions from one social system to another, they may be altered by consensus or general usage within a social system." (Turiel, 1983b; p. 38).

1.2: The moral/ conventional distinction

Developmental psychologists (notably Nucci, Smetana and Turiel) have recently demonstrated that children distinguish between moral and conventional transgressions in their judgements. This distinction can be seen both in the individual's **criterion judgements** and in their **justification categories** (using the terminology of Turiel, 1983a). According to Turiel, criterion judgements define the identifying features of a domain of knowledge. An example criterion judgement is the judgement, by the individual, of a transgression's *rule contingency*; i.e., whether the transgression would be permissable if there was no rule prohibiting the act. Conventional transgressions are defined in the individual's judgements by their rule contingency; subjects judge that if there is no rule prohibiting a conventional transgression, the act is permissable. Moral rules are defined in the individual's judgements by their independence of rules; subjects judge that even if there is no rule prohibiting a moral transgression, the act is not permissable.

According to Turiel, justification categories "form organized systems of thinking within domains" (Turiel, 1983a; p. 53). These justification categories are elicited by asking subjects about their criterion judgements (e.g., "Why do you consider that act is not permissable?"). They are the various forms of response subjects give in reply to these questions. For example, subjects may state that they thought the act was not permissable for reasons based on justice, harm or rights or elements of social system coordination. Justification categories are, in other words, the subject's consciously accessible theories about his criterion judgements; they are the categories of his criterion judgement meta-knowledge (*see*, Karmiloff-Smith, 1986).

The paradigm that has been most commonly practised involves the subject being exposed to characters committing various moral and conventional transgressions (see, tables 1.1 and 1.2 for details on the transgressions that have been used). Usually, the transgressions are presented to the subjects as stories (e.g., Arsenio & Ford, 1985; Davidson, Turiel, & Black, 1983; Hollos, Leis, & Turiel, 1986; Nucci, 1981; Smetana, 1981; Smetana, Bridgeman, & Turiel, 1983; Tisak & Turiel, 1984; 1988; Turiel, 1983a; Weston & Turiel, 1980), however, subjects have also been interviewed after they have witnessed naturally occurring transgressions (Nucci & Nucci, 1982a; Nucci & Turiel, 1978; Nucci, Turiel, & Encarnacion-Gawrych, 1983). After the transgression has been presented to the subject, he is questioned. Each question reflects one of the criteria that Nucci, Smetana and Turiel have postulated will distinguish moral and conventional rules.

Table 1.1: The moral transgression stimuli that have been used (taken from Turiel *et al.*, 1987).

Event Type	Examples of Event Type
Physical harm and welfare	Hitting, pushing, killing.
Psychological harm	Hurting feelings, name-calling, ridiculing a cripple.
Fairness and rights	Stealing, destroying other people's property, breaking a promise, slavery, turn-taking.
Positive "prosocial" moral behaviours	Donating to charity, sharing, helping others in distress.

I have divided the criterion judgements investigated into three categories: seriousness; modifiability and; relativity (see table 1.3).The seriousness category includes those criterion judgements assessing the seriousness of the transgression (the *permissibility*, *seriousness*,

Table 1.2: The conventional transgression stimuli that have been used (taken from Turiel *et al.*, 1987).

Event Type	Examples of Event Type
School rules	Chewing gum in class; boy entering girl's bathroom; eating in class; talking without raising hand; a child not participating in show and tell; a child not sitting in designated place during story time; a child not placing his/ her belongings in the designated place during story time; talking during nap time; standing during snack; child undressing on the playground; leaving the classroom without permission; coming into class late; not lining up in the schoolyard after recess; talking back to the teacher; leaving assigned seat in classroom.
Forms of address	Calling a teacher by his/ her first name; not addressing a judge as "Your Honour".
Conventions governing attire and appearance	Baseball player does not wear uniform to the game; not wearing school uniform; dressing casually in a business office; public nudity; a boy not combing hair.
Game rules	Hide and seek rules; game rules generated by subject.
Family rules	Not clearing one's dishes from the table after dinner; not saying grace before snack.
Customs (e.g., etiquette)	Not using utensils; drinking soup out of a bowl; swearing; girls wearing a skirt sits with her knees apart; husband adopts wife's surname after marriage.
Sex role conventions	Boy wears beret to keep his hair out of his eyes when playing football; bot wears fingernail polish; girl gets a crew cut; girls wears boy's suit; a boy wants to become a nurse caring for infants when he grows up; a boy wears a dress to school; boy or girl uses toys meant for the opposite sex; boy or girl tries to cross gender occupation boundary; sex role conventions provided by the subject.
Sexual conventions	Premarital sex; masturbation; birth control ("the pill"); homosexuality; divorce; marital sex solely for pleasure.
Religious conventions	Not attending Mass on sunday; eating fifteen minutes before communion; receiving communion without confession; going a year without communion; not attending Mass on Christmas or Easter; ordaining women; day of worship; work on sunday; head covering; baptism; interfaith marriage; women preaching; premarital sex.
Other conventions	Meaning of words; greeting (girl greets her friend by bowing rather than the customary greeting); celebrating holidays and birthdays by sending cards.

importance and *punishment* judgements). The modifiability category includes those criterion judgements assessing whether judgements of the transgression are altered by the rule conditions (the *rule contingency, rule knowledge contingency, authority jurisdiction, generalizability*[1] and *flexibility* judgements). The relativity category includes only the *relativity* judgement. I have distinguished these three categories for empirical (see below) and theoretical reasons. These three categories of criterion judgement must involve different processing requirements. The seriousness criterion judgements require value judgements. Since these are value judgements, the moral/ conventional distinction for these criteria can only be one of degree; usually, moral transgressions are judged more seriously than conventional judgements. The modifiability criterion judgements require counterfactual reasoning. The subject must calculate a situation (e.g., where there is no rule against the transgression) and use this to generate a judgement. At least in older subjects, the moral/ conventional distinction for these criteria should be one of difference; without a rule prohibiting the act, conventional transgressions are judged permissable while moral transgressions are judged not permissable. The relativity criterion judgement ("Would the event be O.K. at home or in another school) involves specific knowledge. For this reason, this criterion judgement will no longer be considered.

The research has indicated that individuals frequently judge moral transgressions as more serious than conventional transgressions. While all of the transgression situations, whether moral or conventional, are generally judged not *permissable*, conventional transgressions are more likely to be judged *permissable* than moral transgressions (Smetana, 1985; 1986; Smetana & Braeges, 1990; Tisak & Turiel, 1988; Weston & Turiel, 1980). In addition, subjects generally state that moral transgressions are more *serious* than conventional transgressions or rank them as more *serious* than conventional transgressions (Nucci, 1981; Smetana, 1981; 1985; Smetana & Braeges, 1990; Smetana *et al.*, 1983; Stoddart & Turiel, 1985). Also, subjects generally judge that it is more *important* to have a rule prohibiting moral transgressions than it is to have one prohibiting conventional transgressions (Tisak & Turiel, 1984; 1988; Turiel, 1983a). In line with this, subjects regard that moral transgressions should be *punished* more than conventional transgressions (Smetana, 1981).

---

[1] The terms *generalizability* and *relativity* have been used interchangeably. Thus, Smetana (1981) terms the question, "Might it be O.K. at another school?", the *relativity* criteria while Smetana and Braeges (1990) term this question the *generalizability* criteria. The issue is further complicated by Turiel (Nisan, 1987; Tisak & Turiel, 1984; Turiel, 1983) who asks a rather different question "Would it be alright to do X in another country if no rule prohibited it?" and terms this the *generalizability* criteria. It is this question that is referred to as the *generalizability* question in table 1.3.

Table 1.3: The criterion judgements that have been investigated.

Category	Criterion Judgement	The Question
Seriousness	To measure the *permissibility* of the act	Is it O.K. to X?
	To measure perceived *seriousness.*	How bad is this transgression?
	To measure the perceived *importance* of having a rule prohibiting the act.	How important is it not to X?
	To measure the degree of *punishment* the transgressor should receive.	Should the transgressor be punished? And, if so, a little or a lot?
Modifiability	To measure the transgression's *rule contingency.*	Would it [the depicted event] be OK if there was no rule about it here?
	To measure the transgression's *rule knowledge contingency.*	Tom's teacher never told him it's wrong to X. Is it still wrong?
	To measure the transgression's *authority jurisdiction.*	Suppose the school principal decided that there was no rule in the school against X. Would it be alright to X?
	To measure the transgression's *generalizability.*	Would it be OK to X in another country if there were no rule prohibiting the act there?
	To measure the *flexibility* of the transgression.	Do you think girls/ boys can X if he did not get into trouble and no-one laughed.
Relativity	To measure the transgression's *relativity* to the context.	Would [the event] be OK at home or in another school? Do you think that there might be another country where people do X?

As regards the modifiability category of criterion judgements, the research indicates that moral transgressions are judged differently from conventional transgressions. As stated above, moral transgressions are judged less *rule contingent* than conventional transgressions (Arsenio & Ford, 1985; Smetana, 1981; 1985; Smetana *et al*, 1983; Smetana *et al*, 1984; Nucci, 1981; Nucci & Nucci, 1982a; Nucci & Turiel, 1978; Weston & Turiel, 1980); individuals state that moral transgressions are not permissable even in the absence of prohibiting rules while conventional

transgressions are judged permissable if there is no rule prohibiting them[2]. Moral transgressionsare also less likely to be judged *contingent on rule knowledge* than conventional transgressions (Smetana & Braeges, 1990); the child is less likely to judge that a moral rule is alright to break than a conventional one. Moral transgressions are less under *authority jurisdiction* (the act would not be permissable even if the teacher says that you can do the act) than conventional transgressions (Laupa & Turiel, 1986; Tisak & Turiel, 1984; 1988; Turiel, 1983a). Subjects judge moral rules as less *alterable* than conventional rules (Miller & Bersoff, 1988; Smetana, 1986); subjects are less likely to believe that a moral, rather than a conventional, rule could be changed. Subjects also judge moral rules as more *generalizable* than conventional rules; the act would not be permissable even if, in another country, there was no rule prohibiting the act (Tisak & Turiel, 1984; Turiel, 1983a).

The fact that moral transgressions are generally considered to be more serious than conventional transgressions could be seen as a confound to the research. The moral/ conventional distinction for the modifiability category of criterion judgements might only reflect the subject's judgement that moral transgressions were more serious. Transgressions judged as serious would be judged unmodifiable (i.e., moral) while less serious transgressions would be judged modifiable (i.e., conventional). All transgressions might be measurable across a single continuum; seriousness. Transgressions at either end of this continuum would be judged differently. The moral/ conventional distinction would exist because most moral transgressions would be at the serious end of the continuum and most conventional transgressions would be at the less serious end of the continuum. However, two studies have demonstrated that this is not the case (Stoddart & Turiel, 1985; Turiel, 1983a). These studies have demonstrated that while items in the moral and conventional domains overlap on the seriousness dimension, judgements of modifiability are distinctly different for each domain.

Turiel (1983a; chapter 4) tested this hypothesis with first and fifth grade children by using the following four transgressions: the theft of five dollars [Moral 1 (M1)]; the theft of an eraser (M2); a boy wearing a dress to school [Conventional 1 (C1)]; a child wearing pyjamas to school (C2). The subjects were then asked: how important it was not to do the transgression (assessing

[2] It is interesting to note that while individuals regard moral transgressions, but not conventional transgressions, as wrong even in the absence of rules, they consider that it is more important to have rules that prohibit moral transgressions than it is to have rules which prohibit conventional transgressions (see above). This paradox is probably the result of inadequately updated meta-knowledge.

the *seriousness* of the transgression); whether, if the school principal decided that there was no rule in that school against the transgression, it would be alright to do the transgression (assessing *authority jurisdiction*) and; whether it would be alright to do the transgression in a country where there was no rule prohibiting it (assessing *generalizability*). First graders considered that it was more important not to wear pyjamas to school than to steal an eraser[3] (it was more important not to C2 than to M2). However, while these subjects stated that it was more important not to C2 than to M2, if authority permitted these transgressions or, if in another country there were no rules against these transgressions, the subjects were much more likely to consider C2 (wearing pyjamas to school), rather than M2 (stealing an eraser), was permissable.

Thirteen year old subjects in the Stoddart and Turiel (1985) study judged the transgression of sex role conventions as seriously as they judged the transgression of moral rules. Despite this, they distinguished moral and (sex role) conventional transgressions in their criterion judgements. The children judged that even if the actor would not get into trouble or be laughed at, moral transgressions were **not** permissable while conventional transgressions and sex role deviations **were** permissable. In addition, these children judged that there might be another country where the act was permitted if the act was a sex role deviation or a conventional transgression but not if the act was a moral transgression.

In addition to the moral/ conventional distinction in subject's criterion judgements, a moral/ conventional distinction is also reflected in the subject's justification categories. Subjects use different categories to justify their judgements about moral and conventional transgressions. Table 1.4 defines these justification categories. Basically, children justify their opinions about moral rules by references to the *victim's welfare* and by *appeals to fairness*. Children are more likely to consider moral rules as *categorically wrong*[4]. Children justify their opinion that conventional transgressions are wrong by references to the *disorder* caused, the *existence of* authority or *rules/ laws prohibiting the act* and *efficiency*. Children are also more likely to consider that conventional transgressions are *impolite*[5].

---

[3] However, the fifth graders judged that it was more important not to do either of the moral transgressions than it was not to do either of the conventional transgressions.

[4] This justification category is used by children aged 7 years and younger (Nucci, 1981).

[5] This justification category is used by subjects is their mid-teens and older (Nucci, 1981).

Table 1.4: The justification categories that have been used.

Justification Category	Category Definition
Victim's welfare	Appeal to the interests of the person other than the actor.
Appeal to Fairness	Act deprives someone of what is rightfully theirs, and/ or is unjust; e.g., "He took a toy that belongs to someone else"
Categorically wrong	Global descriptions of the act as bad, wrong or not nice
Disorder and deviation	Act causes chaos, a mess, or is aberrant; e.g., "Eating with your fingers is sloppy."
Existence of authority	Act is sanctioned by authorities and/ or formal rules; e.g., "The teacher said it's against the rules."
Existence of rules/ customs/ norms	Appeal to personal and family customs as to social customs and traditions.
Efficiency	Act will save time or physical effort; e.g., "If he lines up they'll get in quicker."
Politeness	Act is impolite.

1.3: The development of the moral/ conventional distinction

Section 1.2 described the form of the moral/ conventional distinction. This section will detail the development of this distinction. Two points will be considered: the age at which the child begins to distinguish moral and conventional transgressions (section 1.3.1) and; the features that determine whether the child will judge the act as moral or conventional (section 1.3.2).

1.3.1: At what age does the child begin to distinguish moral and conventional rules?

The moral/ conventional distinction is manifested in the child's reasoning after their third year.[6] At 26 and 34 months, the child does not judge moral and conventional transgressions differently on any of the criteria judgement so far used; *permissibility, seriousness, rule contingency, rule*

[6] Smetana and Braeges (1990) report that 34 month olds judged conventional rules as more *relative* than conventional rules. However, given the claim made above that the relativity question mainly tests general knowledge this finding was not taken as evidence that the moral/ conventional distinction is present at this age.

knowledge contingency (Smetana & Braeges, 1990). At these ages, children judge moral and conventional transgression as *non-permissable* and equally *serious*. In addition, they consider both moral and conventional rules as non-modifiable; i.e., as neither *rule* or *rule knowledge contingent.*

By 39 months children are making the moral/ conventional distinction (Smetana, 1981). Smetana (1981) found that by this age children judged moral transgressions as more *serious* and as deserving more *punishment* than conventional transgressions. She also observed that children at this age judged moral transgressions as less *rule contingent* than conventional transgressions. In fact, she observed that children at this age judged conventional rules to be rule **contingent** while they judged moral transgressions to be rule **independent.** When there was no rule prohibiting the act, children's most frequent response to conventional transgressions was to say that they were permissable. When there was no rule prohibiting the act, children's most frequent response to moral transgressions was to say that the act was not permissable. However, this finding that very young children judge conventional rules as rule contingent and moral transgressions as rule independent was not replicated with slightly older children (Smetana & Braeges, 1990). 42 month olds in the Smetana and Braeges study were only more likely to judge conventional transgressions as rule contingent. Their dominant response, even for conventional transgressions, was to state that all of the transgressions were rule independent; i.e., that even if there was no rule prohibiting the act the act was still not permissable.

With increasing age, there is no good evidence of any change in the form of the moral/ conventional distinction for the seriousness criterion judgements. All transgressions are generally judged not permissable but conventional transgressions are more likely to be judged permissable than moral transgressions. Moral transgressions are generally judged more seriously than conventional transgressions.[7] There is evidence of a change in the form of the moral/ conventional distinction for the modifiability criterion judgements. By 5 or 6 years of age, children are generally judging that conventional rules are rule contingent (Davidson, Turiel & Black, 1983; Nucci, 1981).

[7] Though there is some evidence that at some ages some moral transgressions are judged less seriously than some conventional transgressions (see section 1.2). However, it appears from the Stoddart and Turiel (1985) and Turiel (1983; chapter four) studies that by at least the mid teens all moral transgressions are judged more seriously than all conventional transgressions.

1.3.2: The determinants of a rule's domain

Turiel (1983a) and Davidson, Turiel & Black (1983) have claimed that the child's familiarity with a rule determines whether the rule is judged as moral or conventional. Their claim is based on the evidence of one study; that of Davidson *et al.*, (1983). Davidson *et al.* presented children with four transgression situations: a familiar moral (FM); an unfamiliar moral (UM); a familiar conventional (FC) and; an unfamiliar conventional (UC). Familiarity was determined by presenting the subject with the transgression situation and asking them whether they knew someone who had experienced the transgression, or they had participated in the transgression, or they had clear familiarity with the transgression from some other source. Familiar transgressions were those where the subject knew someone who had experienced the transgression or they, themselves, had participated in the transgression, or they had clear familiarity with the transgression. Unfamiliar transgressions were those where the subject did not know anyone who had personally experienced the transgression, had not observed the transgression and had no obvious familiarity with the transgression. The most commonly used familiar and unfamiliar transgressions are listed below:

FM: A boy pushes another child off a slide.
A student steals a dollar from another child.

UM: An emergency room surgeon leaves his post even though there is no-one to relieve him.
An employee embezzles money from the company he works for.

FC: A child eats dinner with her fingers.

UC: A couple marry and the husband adopts the wife's surname.
A child addresses his teacher by his first name.

Five forms of criterion judgements were investigated in the study. These were: overall evaluation (e.g., "Was the act O.K.?"), subordinate jurisdiction over rule (e.g., "Could the children get the rule changed?"), Acceptance of authority (e.g., "Can an authority figure abolish this rule?"), punishment orientation (e.g., "Is the act all right if punishment is expected?"), acceptance of rule change (e.g., "Is it alright to change this rule?"), understanding of subjective

responsibility (e.g., "Is the act wrong if well intentioned?").

In line with previous research, the results indicated that the subjects: (1) evaluated the moral transgressions more seriously than the conventional transgressions; (2) accepted less authority power with respect to moral issues than with respect to conventional issues; (3) considered that the expectation of punishment influenced the evaluation of a conventional transgression more than the evaluation of a moral transgression; (4) judged changes in rules more legitimate if the rule was conventional rather than if it was moral; (5) did not think that either type of rules was under the jurisdiction of children. As regards the familiarity/ unfamiliarity issue, Turiel observed that while subjects judged FM items more seriously than FC on the punishment scale, subjects did not distinguish between UM and UC items on this scale. However, it should be noted that subjects judged UM issues more seriously than UC issues, they judged that there should be less authority power over UM than UC issues and that change was less legitimate for a UM rule than a UC rule. Also, there is a potential confound to Turiel's conclusion; the familiar moral items involved clear victims (the person getting hit, the victim of the theft) while the unfamiliar moral items did not (in one story the person embezzled from a company, not a person, and in another the actor was being irresponsible; there would only be a victim to the doctor's action if a patient was admitted to the emergency room when the doctor was absent). It is thus not clear from this study whether the familiarity of a rule determines its domain or whether the obviousness of the intrinsic consequences of the transgression determines its domain. A study by Smetana (1985) indicates the latter interpretation.

Smetana (1985) investigated the role of the perceived consequences of a transgression in determining whether that transgression is judged as moral or as conventional. She compared children's judgements of familiar moral and conventional transgressions with their judgements of unspecified transgressions. These unspecified transgressions were referred t by nonsense words. For example, the child was told that: "Mary did serat ... ". The unspecified transgressions differed in their consequences according to two criteria, which matched differences between familiar, conventional and moral transgressions. The first criterion listed consistency of the consequence. Thus, the familiar/conventional transgressions are contextual in their application, applying only at school or only at home, while the familiar/moral transgressions were applicable in both places. The second criterion was the type of response. The familiar/conventional transgression results in the adult appealing to rules. This was matched by an undefined contextualised transgression. The moral transgressions give rise to two kinds of

responses; the child victim cries and the adult tells the child perpetrator to stop doing the act (referred to as "adult commands" in the table). The final three unspecific transgressions, then, were distinguished by having one or both of these types of response in a generalizable context. The six conditions used are summarized in Table 1.5.

Table 1.5: The six stimulus conditions used in the Smetana (1985) study.

Condition	Event	Consistency of consequence	Respondent/ response types
Conventional	Talking during naptime Standing during snack	Contextual	Adult - appeal to rules
Undefined	Unspecified	Contextual	Adult - appeal to rules
Moral	Stealing cookies Hurting feelings	Generalizable	Adult commands/ child cries
Child + adult response	Unspecified	Generalizable	Adult commands/ child cries
Child response	Unspecified	Generalizable	Child cries
Adult response	Unspecified	Generalizable	Adult commands

Smetana compared responses to the familiar/conventional transgression to the undefined one which matched it in consequences. The two judgements did not differ significantly on any of the criteria on judgements studied except *seriousness*; the children judged that the familiar conventional transgression was more serious than the undefined transgression. Judgements of the familiar/ moral transgression were compared with the three unspecified conditions where the outcomes were the same. First of all we can take the unspecified transgression which just led to the adult response of telling the child to stop doing it. This transgression was seen by the children as less serious, more authority dependent and more rule contingent than the familiar/ moral transgression. In contrast to this, the unspecified transgressions that resulted in the child crying were judged in exactly the same way as the familiar/moral transgression. That is, they were judged as being just as serious, and just as free from authority or rules as the moral transgressions. One can conclude from this that the children judged those transgressions which resulted in clearly identifiable victims as moral transgressions even if the actual act was unspecified.

The comparison of conventional and moral transgressions showed that the familiar/moral

transgression was more serious, less authority dependent and less rule contingent than the conventional transgression. The judgements of the unspecified events enabled us to sort out which component is the crucial one. In summary then, the Smetana (1985) study indicates that it is the presence of harmful consequences to another, in this case, specifically, a child crying, which determines that the transgressions will be judged as moral.

1.4: Criticisms of the moral/ conventional distinction

Several writers have critically approached the notion of a moral/ conventional distinction (e.g., Gabennesch, 1990; Rest, 1983; Shweder, Mahaptra & Miller, 1987). Their criticisms have either focused on the theoretical basis of the distinction (e.g., Rest, 1983; Shweder *et al.*, 1987) or on the data (Gabennesch, 1990; Shweder *et al.*, 1987). The theoretical attack on the moral/ convention distinction will be reviewed in section 1.4.1. Work that has been cited as indicating that the child does not distinguish between moral and conventional rules will be reviewed in section 1.4.2.

1.4.1: A theoretical attack on the moral/ conventional distinction

Rest (1983) and Shweder (e.g., Shweder, Mahaptra, & Miller, 1987) have both criticised the moral/ conventional distinction. We will see that their criticism has focused on the defining issues of the moral/ conventional distinction; that moral rules are associated with the welfare of other people and that conventional rules are associated with social organization. It will then be argued that their attack is invalid because of this focus on definition; issues of definition are philosophical level issues which cannot be used to attack a position at the psychological level. It is irrelevant whether or not moral and conventional transgressions can be distinguished at some abstract philosophical level. It is only relevant to know whether they are distinguished by subjects. However, Rest has suggested that:

> "this distinction is not adequate because concepts of social organization are logically implicated in both morality and social convention - concepts of justice provide standards by which to judge social organization. How could there be justice or morality *without* social organization?" (Rest, 1983; p. 609).

In addition to Rest's claim that morality cannot be distinguished from convention because they both involve social organization, Rest and Shweder have claimed that the two cannot be

distinguished because they can both have implications for the welfare of others. For example, Rest states:

> "a person's welfare is also affected indirectly by general social structures and social arrangements... conventions ... are institutionalized practices ... designed to further human welfare at least indirectly; therefore we cannot relegate these social practices to a domain separate and independent of morality" (Rest, 1983; p. 610).

and Shweder claims:

> "the Turiel, Nucci, and Smetana theory ... underplays the way ritual observances and customary practices involving food, sex, dress, the exchange of greetings, and terms of address may be linked through social meanings to mandatory moral principles like harm, justice, and natural law" (Shweder *et al.*, 1987; p. 29).

Rest points out that breaking the conventional rule of driving on the opposite side of the road to that proscribed by the culture will have welfare implications. Shweder points out that the violation of a dress code in certain cultures may have emotional consequences for observers.

Of course, these criticisms have no bearing on the claims of Turiel and his colleagues. They stem from a confusion of the philosophical and the psychological levels. Definitional issues, such as the ones causing the problems here, are philosophical issues. Rest and Shweder are perfectly correct to point out that moral rules cannot necessarily be distinguished from conventional rules at a philosophical level. There can never be any classification of acts as moral and conventional at the philosophical level. Any act may be considered to result in victims and therefore be *judged* moral. Any act may be considered as not resulting in victims (by reasoning that the person deserved their pain) and therefore be *judged* conventional. But this is a philosophical issue. It has no bearing on the Turiel, Nucci and Smetana position which is at the psychological level. Their findings indicate that individuals **process** rules differently according to whether they *judge* that breaking that rule inflicts other person pain or not. The theoretical attack on the moral/ conventional distinction is thus invalid.

1.4.2: Work cited as disproving the moral/ conventional distinction

The work that has been cited as indicating that the child does not distinguish between moral and conventional rules can be divided into three categories: (1) work indicating moral transgressions

are not always considered to be more serious than conventional transgressions (section 1.4.2.1); (2) work indicating that young children do not always treat conventional rules as conventional (section 1.4.2.2) and; (3) cross-cultural work (section 1.4.2.3).

1.4.2.1: Work indicating that moral transgressions are not always considered more serious than conventional transgressions

Gabennesch (1990) cites Damon (1977) who found that 63% of 4- and 5- year olds and 40% of 6- year olds considered poor table manners to be as wrong as stealing and Shantz (1982) who reported that the majority of her 6- and 7-year-old subjects did not consistently rank conventional violations as less serious than moral conventions. Similarly, Shweder *et al.*, (1987) cite Murdock (1980) who conducted a cross-cultural survey about the causes of illness. Shweder *et al.*, (1987) argue, not unreasonably, that if a rule violation is thought to make you sick the rule must be considered as important within the culture. They then notice that conventional rule violations (e.g., violations of food and sex taboos) are more associated cross-culturally with illness, and therefore as serious, (and therefore important) as moral rules. They suggest that since transgressions of these taboos are considered as serious as moral transgressions these taboos cannot be considered nonmoral rules.

Of course, both these attacks on the validity of the moral/ conventional distinction only demonstrate the profound confusion of their authors. These authors are treating *seriousness* as one of the definitional criteria for morality. However, this is inappropriate. As stated in section 1.2, the criterion judgements that have been used can be divided into two categories: *seriousness* judgements and *modifiability* judgements. While moral transgressions are generally judged more seriously than conventional transgressions they are not defined by their seriousness; some conventional transgressions are judged more serious than some moral transgressions at some ages (Stoddart & Turiel, 1985; Turiel, 1983a; chapter four). Moral transgressions are defined, psychologically, by their judged immodifiability.

1.4.2.2: Work indicating that young children do not always treat conventional rules as conventional

Gabennesch (1990) cites several studies that indicated conventional rules are not differentiated from moral rules; more specifically that they are not more alterable and relative. Unfortunately,

most of his citations are inappropriate and indicate that he did not understand the data (*see* Helwig, Tisak, & Turiel, (1990) for detailed demonstrations of his ignorance). However, two of his citations (Carter & Patterson, 1982; Komatsu & Galotti, 1986) do bear closer scrutiny.

Komatsu and Galotti (1986; study 1) found that the first graders did not regard game rules, the school calendar, the names of objects and activities, and driving customs as alterable. However, third and fifth graders and the adult subjects did regard these conventions as alterable. Carter and Patterson (1982) reported that half of their second and fourth graders did not attribute flexibility and cultural relativity to the norm prohibiting eating with one's fingers. However, sixth and eighth graders did attribute flexibility and cultural relativity to this norm.

Thus, Gabennesch is right. At very young ages, children do not always consider conventional rules as alterable and relative. However, this is irrelevant. In section 1.3.1, I stated on the basis of the findings of Davidson *et al.* (1983) and Nucci (1981) that conventional transgressions were judged modifiable on the modifiable criterion judgements from the age of 6 or 7. The Komatsu and Galotti study reported here suggests that this may actually be achieved later.[8] The Carter and Patterson study suggests later still. However, the important fact is not when conventional transgressions are judged modifiable but if they are judged modifiable. The older children in both of these studies judged all of the conventional transgressions modifiable. Thus, these studies cannot be used as evidence against the existence of the moral/ conventional distinction.

1.4.2.3: Cross-cultural work

Six studies have explored the cross-cultural validity of the moral/ conventional distinction (Edwards, 1987; Hollos, Leis, & Turiel, 1986; Nisan, 1987; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Shweder, Mahaptra & Miller, 1987; Song, Smetana, & Kim, 1987). These studies have followed three different paradigms: Edwards (1987) investigated behavioral reactions to moral and conventional transgressions; Shweder *et al* (1987) devised their own paradigm and; the other four studies followed the classic paradigm described in section 1.2.

Edwards (1987) analyzed 109 episodes of the behaviour of a Luo-speaking community in the south Nyanza district of Kenya. These episodes had been collected by another researcher.

[8] In the second Komatsu and Galotti study (1986; study 2), however, first graders did regard these conventional transgressions as alterable and relative.

Edwards provided examples of the responses of individuals to various types of transgression but attempted no statistical analysis of the data. Two of her social situation categories (suggestions related to meeting the needs of others and discussions about task assignment) are not even transgression situations. However, in contrast to studies of the social interactions following transgression in Western cultures (see section 1.6.1.1), Edwards reports that the behavioral responses of adults and peers to moral and conventional transgressions are the same. According to Edwards' data, the only form of response of these people to any form of transgression is power assertive (physical punishment is provided or threatened). This surprising finding must be treated with some caution, however, because of the totally unscientific way in which she has presented her data. All the data is presented as anecdotes on which no statistical tests or any other forms of analysis have been conducted. However, when considering the data it should be noted that all it demonstrates is that this community do not distinguish moral and conventional transgressions in their *behaviour*. A lack of a behaviourial difference does not necessarily indicate the lack of a cognitive difference. This data does not show that these individuals do not *judge* moral and conventional transgressions differently.

Shweder *et al.*, (1987) investigated the judgements of two Indian samples (a Brahman and an "Untouchable" population) and an American population. Thirty-nine transgression situations, representing a range of family life and social practices, were presented to each subject. Analysis of the thirty-nine situations reveals the first of the flaws that make this study worthless. All the situations involving victim harm involved mitigating circumstances. For example, in one a husband beat his wife *for disobeying him.* In another, a child stole *on his father's orders.* Of the nine situations spotted by this author which directly involve victims, in only two is the transgression unprovoked (a classmate tears up a girl's drawing, a man kicks a sleeping dog). These two situations are not analyzed separately.

Shweder *et al* (1987) asked his subjects several questions in order to assess the contingency on other person knowledge, relativity and alterability of the rule that was broken in the transgression story. Here Shweder *et al.*, confound their data again. They use three questions to elicit the subject's judgement about the rule's relativity and alterability. However, the questions that were asked concerned the rule the subject invoked when reasoning about the transgression. For example, one transgression situation involved a widow eating fish. Interviews for subjects might be as follows (these examples are taken from Shweder *et al.*, 1987; p. 43-44):

Subject one (Indian population):

1) Is the widow's behaviour wrong? (Yes. Widows should not eat fish, meat, onions or garlic or any "hot" foods. They must restrict their diet to "cool" foods, rice, dhal, ghee, vegetables.] ....

5) Would it be best if everyone in the world followed the rule that widows should not eat fish? (That would be best. A widow's devotion to her deceased husband - who should be treated like a god. She will offend his spirit if she eats fish.]

Subject two (American population):

1) Is the widow's behaviour wrong? (No. She can eat fish if she wants to] ....

5) Would it be best if everyone followed the rule *that it is alright for a widow to eat fish if she wants to*? [Yes. People should be free to eat fish if they want to. Everyone has that right.]

From these examples it can be seen that subjects were asked totally different questions according to how they had responded initially. Subject one was asked about the rule that a widow should not eat fish, subject two was asked about the rule that a widow should be able to eat fish is she wants to. Asking subject two about the rule that a widow should be able to eat fish if she wants to is a confound to the purported status of the task which is to discover how individuals judge the rule that widows should not eat fish. In my opinion this second confound is so enormous that it makes any further review of this paper purposeless.

Fortunately, four unconfounded cross-cultural studies have been undertaken (Hollos, Leis, & Turiel, 1986; Nisan, 1987; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Song, Smetana, & Kim, 1987). All four followed the classic paradigm described in section 1.2. In three of the four studies the transgression situation was described to the subject (Hollos *et al.*, 1986; Nisan, 1987; Song *et al.*, 1987), in the fourth (Nucci *et al*, 1983), the subject's judgements of naturally occurring transgression situations they had witnessed, were investigated. These studies were conducted in Nigeria (Hollos *et al*, 1986), Israel (Nisan, 1987), the Virgin Islands (Nucci *et al.*, (1983) and Korea (Song *et al.*, 1987). Hollos *et al* (1986) observed that rural Nigerians (aged

8-11, 12-14 and 15-18 years) judged a conventional rule (a form of greeting) as more *authority dependent* (the subject was more likely to think that it was alright for an actor to do an act if an authority figure directed them to do the act) and more *alterable* than the moral rule (not stealing). Nucci *et al* (1983) found that 3-5 year old (mostly rural) Virgin Islanders considered conventional rules more *rule contingent, generalizable* (it would be alright for another country not to have the rule) and *act relative* (if another country does not have the rule against an act it would be alright for people to do the act) than moral rules. Song *et al.*, (1987) found that urban Koreans (mean ages = 5.37, 9.43, 12.40, 15.10 and 18.13) judged conventional rules more *relative* and less *rule contingent* than moral rules.

Nisan (1987) investigated the judgements of a group of Israeli Jews (urban and kibbutzim) and a group of Israeli Moslems. He found that the reduction in seriousness of a transgression following its permission by law was significantly greater if the rule was conventional than if it was moral. Nisan actually suggests that his data is problematic for the moral/ convention distinction position. His argument is based on the fact that the two groups differ in their responding (for example the Israeli Moslems treated all the transgressions as more serious than the Israeli Jews). He suggests that this means that the moral/ conventional distinction is simply a product of cultural pressures. However, while this finding is interesting (it does, indeed, indicate a cultural influence in the formation of the moral/ conventional distinction) it neither disproves the moral/ conventional distinction or indicates that the distinction is simply a product of cultural pressures.

Three of the four studies report the justification categories of the subjects. Moral items were justified primarily in terms of other's welfare (Hollos *et al.*, 1986; Nisan, 1987;[9] Song *et al.*, 1987), obligation (Song *et al.*, 1987), fairness (Nisan, 1987; Song *et al.*, 1987) and by appeal to law (Hollos *et al.*, 1986) or religion (Nisan, 1987; the Israeli Moslem population). Conventional items were justified in terms of authority (Hollos *et al.*, 1986; Song *et al.*, 1987), sanctions (Song *et al.*, 1987), social nonconformity (Song *et al.*, 1987), social coordination (Song *et al.*, 1987), customs and traditions (Hollos *et al.*, 1986), and prudential (Song *et al.*, 1987), pragmatic (Song *et al.*, 1987), and personal reasons (Song *et al.*, 1987).

[9] Nisan (1987) did not examine justification category by domain since he was determined to demonstrate a difference between the Jewish and Moslem Israelis in their reasoning. Since this difference has no bearing on the moral/ conventional distinction, I have not reported it. The statistical analyses on his data were conducted by this author.

1.5: The importance of the moral/ conventional distinction; theories of the development of morality (1)

The observation that individuals distinguish between moral and conventional rules from an early age is important because the theories of the development of morality that were dominant before this observation would not predict its existence. Piaget (1932), and those who attempted to extend his formulation (e.g., Kohlberg, 1969; Rest, 1983; Youniss, 1980), proposed sequential models of moral development where morality was generated after, or from, convention. Three of these older formulations will be very briefly described below (Rest's formulation will not be described because of its similarity to Kohlberg's).

1.5.1: Piaget's model

Piaget (1932) used a variety of tasks to identify two qualitatively different ways of evaluating rules which he termed **moral realism** and **autonomous morality**. Piaget termed these two ways of evaluating not stages but processes (Piaget, 1932; p. 120). Up to age 10, the two types of answers occur side by side, even within the same child. However, with increasing age, the frequency of moral realist responses is reduced, though, even in adulthood, these responses may be found under certain conditions (Hoffman, 1970). **Moral realism** involves the child evaluating another's actions in terms of their consequences, the extent to which it conforms to established rules, and whether it is punished. The child, when judging the actions of others, tends to ignore the intention of the protagonist.[10] Rules are viewed as sacred and unalterable and, sometimes, of divine origin. The child frequently believes in "imminent justice"; that the violation of a social norm will be followed by physical accidents or misfortunes willed by God, etc. **Autonomous morality** (which develops from around eight) involves the child viewing rules as established and maintained through reciprocal social agreement and thus subject to modification according to human needs. The judgements of the child regarding the action of others stresses the intention of the protagonist rather than the material consequences of the action. The child considers that punishment should be reciprocally related to the misdeed (e.g., through restitution) rather than as pain administered by authority.

[10] This is not to say that the child is unable to process information about the intention of the protagonist. Piaget (1932) was clear that the child does use this form of information in other circumstances. However, the moral realist child does not use this information when judging the actions of others.

Piaget views these two methods of evaluation as the product of two different forms of socialisation experience. Piaget proposes that moral realism is the result of adult constraint. Piaget (1932; p. 131) states that rules are imposed by the revered adult and because of this acquire their sacred status. Piaget proposes that autonomous morality is a result of socialisation experiences with peers. As the child grows older, he begins to think of adults and older children as equals, diminishing his unilateral respect for them, and allowing him to engage in discussion about various rules. Exposure to this discussion makes a moral realist conception of rules untenable. The child may consider some rules to be the product of cooperation and amenable to change by mutual consent. When the child is discussing with his peers, he may come into contact with conflicting moral views which may be evaluated using his developing ability for deductive processing (Piaget, 1932; p. 37).

In summary, Piaget believes that there are two ways of evaluating rules; moral realism and autonomous morality. These two methods of evaluation are seen as products of the child's socialisation experiences. Moral realism is seen as a result of direct communications from the adult while autonomous morality is seen as a result of socialisation experiences with peers.

1.5.2: Youniss's Social Construction Theory

Youniss's (1981, 1987) position should be considered, as Youniss does himself, as a reinterpretation of the work of Piaget. His main contribution has been to stress the role of communication, as opposed to logical deduction, in the development of morality. Youniss states that there are two types of communication situation which may result in the acquisition of moral codes. One type of communication situation involves the unilateral transmission of adult knowledge to children. The child makes only a minimal contribution to the dialogue, he/ she simply acquires the standard advocated by the parent. Youniss claims that this form of communication cannot be the basis of mature morality partly because the child will not view himself as an agent involved in the code construction.

The second type of communication situation occurs where there is confrontation between the multiplicity of views the child encounters through interaction with his peers. Youniss claims that the child attempts to reach a solution to the problem of the variety of viewpoints by consensus-seeking discourse. Through this completely unspecified process, children are assumed to discover that by working together they can produce results that are mutually satisfactory and

mutually understood. Youniss assumes that laws are discovered through this discourse. Any child can suggest a rule but he must persuade his peers, through discourse, of its worthiness. According to Youniss, this form of communication is the basis of mature morality because the child does view himself as an agent involved in the construction of the rule.

Youniss incorporates an affective component into his theory. Affect in this framework is seen to be the commitment people make to work for the common good (Youniss, 1981). Youniss assumes that this affect is generated by the cooperation between peers that is necessary for the generation of rules. The actual process of the generation of this affect is unspecified. However, potential accounts of affect generation in relationship situations are available (e.g., Berscheid, 1982; 1983). This affect is assumed to bind people together and is thus, presumably, rewarding. The individual is thus assumed to be motivated to follow the societal rules so that his binding with the other members of his society remains intact.

In summary, Youniss argues that there are two ways in which the child may acquire moral rules. The first was also described by Piaget (1932); the rule is supplied to the child as a unit which should, in future, direct his behaviour. The second is similar to Piaget's account of the development of autonomous morality except that Youniss stresses inter-child discourse as a situation for the generation of rules and not deductive reasoning. Youniss, unlike Piaget, also considers an affective component. Affect is generated through the cooperation necessary for discourse and is rewarding, rules, as prescribed instances of cooperation, are adhered to so that this rewarding affect will not be lost.

1.5.3: Kohlberg's account

Kohlberg and colleagues (Colby, Kohlberg, & Kaufman, 1989; Kohlberg, 1969; 1971; 1976; 1981; Kohlberg & Kramer, 1969) proposes that moral development, or, at least, the development of justice reasoning (Kohlberg, 1981; 1984), occurs in a series of six qualitatively distinct stages. Kohlberg (Colby, Kohlberg & Kaufman, 1989) define stages as structures, modes of thinking, which serve the same function at various points of development. These stages are viewed as invariant in sequence. Each stage is hypothesized as forming a "structural whole"; a given stage response does not represent a specific response determined by knowledge and familiarity with that task or tasks similar to it; rather, it represents an underlying thought organisation" (Colby, Kohlberg & Kaufman, 1989; p. 6). It is also proposed that stages integrate the structures found

Table 1.6: Kohlberg's six stages of moral judgement (taken from Kohlberg, 1976).

Level and stage	Contents of stage		
	What is right	**Reasons for doing right**	**Sociomoral perspective of stage**
Level 1: Preconventional: Stage 1. Heteronomous morality.	To avoid breaking rules backed by punishment, obedience for its own sake, and avoiding physical damage to persons and property.	Avoidance of punishment and the superior power of authorities.	Egocentric point of view. Doesn't consider that interests of others or recognize that they differ from the actor's, doesn't relate two points of view. Actions are considered physically rather than in terms of psychological interests of others. Confusion of authority's perspective with one's own.
Stage 2. Individualism, instrumental purpose, and exchange.	Following rules only when it is to meet one's own interests and needs and letting others do the same. Right is also what's fair, what's an equal exchange, a deal, an agreement.	To serve one's own needs or interests in a world where you have to recognize that other people have their own interests too.	Concrete individualistic perspective. Aware that everybody has his own interests to pursue and these conflict, so that right is relative (in the concrete individualistic sense).
Level 2: Conventional: Stage 3. Mutual interpersonal expectations, relationships, and interpersonal conformity.	Living up to what is expected by people close to you or what people generally expect of people in your role as son, brother, friend etc. "Being good" is important and means having good motives, showing concern about others. It also means keeping mutual relationships, such as trust, loyalty, respect and gratitude.	The need to be a good person in your own eyes and those of others. Your caring for others. Belief in the Golden rule. Desire to maintain rules and authority which support stereotypical good behaviour.	Perspective of the individual is relationships with other individuals. Aware of shared feelings, agreements, and expectations which take primacy over individual interests. Relates points of view through the concrete Golden Rule, putting yourself in the other guy's shoes. Does not yet consider generalized system perspective.
Stage 4. Social system and conscience.	Fulfilling the actual duties to which you have agreed. Laws are to be upheld except in extreme cases where they conflict with other fixed social duties. Right is also contributing to society, the groups, or institution.	To keep the institution going as a whole, to avoid the breakdown of the system "if everyone does it", or make the imperative of conscience to meet one's defined obligations.	Differentiates societal point of view from interpersonal agreement or motives. Takes the point of view of the system that defines roles and rules. Considers individual relations in terms of place in the system.
Level 3: Postconventional or principled: Stage 5. Social contract or utility and individual rights.	Being aware that people hold a variety of values and opinions, that most values and rules are relative to your group. These relative rules should usually be upheld, however, in the interest of impartiality and because they are the social contract. Some nonrelative values and rights like life and liberty, however, must be upheld in any society and regardless of the majority opinion.	A sense of obligation to law because of one's social contract to make and abide by laws for the welfare of all and for the protection of all people's rights. A feeling of contractual commitment, freely entered upon, to family, friendship, trust and work obligations. Concern that laws and duties be based on rational calculation of overall utility, "the greatest good for the greatest number."	Prior-to-society perspective. Perspective of a rational individual awareness of values and rights prior to social attachments and contracts. Integrates perspectives by formal mechanisms of agreement, contract, objective impartiality, and due process. Considers moral and legal points of view; recognizes that they sometimes conflict and finds it difficult to integrate them.
Stage 6. Universal ethical principles.	Following self-chosen ethical principles. Particular laws or social agreements are usually valid because they rest on such principles. When laws violate these principles, one acts in accordance with the principle. Principles are universal principles of justice: the equality of human rights and respect for the dignity of human beings as individual persons.	The belief as a rational person in the validity of universal moral principles, and a sense of personal commitment to them.	Perspective of a moral point of view from which social arrangements derive. Perspective is that of any rational individual recognizing the nature of morality or the fact that persons are ends in themselves and must be treated as such.

at lower stages. These stages are described in Table 1.6. Kohlberg proposes that these stages are constructed by the child by thinking about and acting on the world. They are not the product of socialisation.

Kohlberg's six stages are based on extensive case analyses of boys ranging from 10 to 16 years of age (Kohlberg, 1969; Kohlberg and Kramer, 1969). Data is collected through interview. The subject is given nine hypothetical moral dilemmas in which there is conflict between an act of obedience to laws, rules, or commands of authority and an act which is congruent with the needs or welfare of other persons. The subject must decide which act to perform. The subject must then state why they made that choice. The subject is then scored on the quality of the explanation of their choice. This scoring procedure has changed several times (Kohlberg, 1969; Kohlberg and Kramer, 1969; Kohlberg, Levine & Hewer, 1983; Colby, Kohlberg & Kaufman, 1989).

1.5.4: Conclusions

It should be clear from the brief descriptions just provided that none of the three formulations would predict that the child distinguished moral and conventional rules from the age of three. At this age the child, according to all three writers, has *learnt* all his rules from interactions with adults. Effectively, at this age, according to these formulations, the child's moral knowledge has been imposed on him by the adult (according to all three formulations, it is only when the child is older that he is actually involved in the rule construction process). Such positions cannot account for the child judging different rules differently.

In addition, it is possible to draw some conclusions about some of the distinctions that can be drawn both within and between theories of moral development. These early formulations (at least those of Piaget and Youniss) distinguish between rules that are learnt, that are imposed on the child from adult interaction (giving rise to moral realism in Piaget's formulation) and rules that are constructed by the child (resulting in autonomous morality). A related distinction is the implicit distinction within these theories between moral and conventional rules; moral rules are self-constructed (by the older child), conventional rules are imposed by the individual's cultural surroundings. Alternatively, moral rules, as self-constructions, must have their own "logic" - they are the product of reasoned argument within the individual. Conventional rules on the other hand are not rational; there is no "logical" reason for their existence *in that particular form*, they

exist only because of societal requirements.

A distinction that can be made between theories is the distinction between those theories that are more interested in the development of the child's processing capacities (e.g., Piaget was interested in the two different processing styles of moral realism and autonomous morality and in how intention information was added to the calculation process determining punishment for action) and those interested in changes in the child's meta-knowledge (e.g., Kohlberg).

1.6: Accounts of the moral/ conventional distinction; theories of the development of morality (2)

As was reported above, the older formulations cannot account for the distinction made by the child between moral and conventional rules. In this section, I will describe two formulations that specifically attempt to account for this distinction.

1.6.1: Constructivist Theory

Turiel, Smetana and Nucci have all formulated broadly similar accounts of the origin of the moral/ conventional distinction (Nucci, 1982; Smetana, 1983; Turiel, 1983a; Turiel & Davidson, 1986; Turiel & Smetana, 1984). These accounts have two components: a description of the origin of the distinction (Nucci, 1982; Smetana, 1983; Turiel, 1983a) and; a description of the cognitive structures that underlie this distinction in the mind (Turiel, 1983a; Turiel & Davidson, 1986; Turiel & Smetana, 1984). These two components will be explored in sections 1.6.1.1 and 1.6.1.2 respectively.

1.6.1.1: A description of the origin of the moral/ conventional distinction

Nucci, Smetana and Turiel assume, like Piaget, that the child's concepts are constructed through his actions upon the environment. They argue that the types of conceptual structures constructed by the child are influenced by the nature of the environment. Different types of events will give rise to different types of concepts (Smetana, 1983; Turiel, 1977; 1978; 1983a). Turiel and Smetana assume that the differences in moral and conventional transgression events (see section 1.6.1.2) result in the different conceptual structures of the moral and conventional domain.

There is some evidence demonstrating the social interactions with peers are important for the formation of the moral/ conventional distinction. Siegal and Storey (1985) compared two groups of four year old children. One of these groups had been attending kindergarten for 18 months or more, the other less than 3 months. While both groups thought that moral transgressions were more serious and should be punished more than conventional transgressions, the distinction made by the short-stay kindergarten children was far smaller. Also, while the long stay kindergarten children considered conventional rules more *rule contingent* than conventional rules, the short stay kindergarten children did not.

It should be clear, however, that Turiel and Smetana are not claiming that these differences in event types "impose" conceptual differences in the mind. Turiel and Smetana claim that the individual *constructs* conceptual frameworks out of his interactions with the environment. Turiel and Smetana propose that the differences in the events surrounding moral and conventional transgressions result in the construction of different types of concepts. Turiel (especially, Turiel, 1977; p. 101) claims that the individual has a variety of methods available to him for this construction of concepts; observation, communication, imitation and role-taking. It should be noted, however, that he does not detail how the child utilizes the various methods for the construction of the moral/ conventional distinction. However, some speculations can be made in this direction. All the methods Turiel (1977) mentions apart from role taking (i.e., observation, communication and imitation) are methods by which information is copied from one individual to another. Role taking is the only one of these methods that actually involves the construction of information by the child. Thus, if Turiel (1977) is to claim that the moral/ conventional distinction is not a direct product of cultural transmission (that the culture does not "impose" the domain structure on the child), he must consider that role taking is the principal method for the construction of the cognitive structures that give rise to the moral/ conventional distinction.

Turiel (1983a) assumes two additional methods are involved in the construction of the moral/ conventional distinction: manipulations of personal past experience and; counter-factual reasoning. He attempts to demonstrate the influence of these two methods through examples. According to Turiel (1983a), being the victim of a moral transgression should immediately result in a sense of the undesirability of the event because of the unpleasantness of being a victim. The child will develop a prescription against this unpleasant event. Turiel also claims (Turiel, 1983a; p, 43) that the child will arrive at "judgements of moral necessity" through comparison

of the performance of the act itself with its opposite. If the constructed consequences of its non-occurrence (there is no victim) are judged to be more "desirable" than the consequences of its occurrence (the victim is harmed), then inferences will be made regarding how people should act in these circumstances. According to Turiel, this same process when applied to conventional transgressions, will not result in automatic prescriptions. Turiel argues that comparison of a conventional act with its opposite will not result in one situation being obviously superior to the other. According to Turiel (1983a), it is social organizational factors, such as consensus, rules, and authority that provide meaning to conventional prescriptions.

It is questionable, however, whether either of these methods could result in the moral/ conventional distinction. Taking the experience of being a victim: why should the child, following this form of interaction, generate general prescriptions against such action as committing harm. It is understandable that the child might generate prescriptions against allowing himself to be a victim but why would he be interested in the welfare of others? Also, taking the counterfactual reasoning account, there is no reason, given Turiel's account, why an individual should consider the constructed consequences of a moral transgression's non-occurrence (there is no victim) more "desirable" than the consequences of its occurrence (the victim is harmed). Turiel does not specify a mechanism that would result in individuals considering harm to another undesirable. In addition, and conversely, it would seem that comparison of many conventional transgressions with their opposites would result in a prescription against the transgression. For example, comparison of talking in class with not talking in class should result in a prescription against talking in class if the individual desires to work.

Then there is the problem that the accounts Turiel gives are intended to be for the development of moral prescriptions. They explain why moral transgressions are judged more seriously than conventional transgressions, they explain why moral transgressions are more likely to be judged unmodifiable than conventional transgressions. They do not explain why conventional transgressions come to be judged modifiable. Yet this is the developmental history (see section 1.3.1). Until the end of the third year, the child judges moral and conventional rules unmodifiable (Smetana, 1981; Smetana & Braeges, 1990). At 42 months, the child still judges moral and conventional transgressions as unmodifiable but he is more likely to judge conventional transgressions than moral transgressions modifiable (Smetana & Braeges, 1990). It is not until 6 or 7 years, that the child regularly judges conventional transgressions modifiable

(Davidson *et al.*, 1983; Nucci, 1981). There is nothing in Turiel's account which states why conventional transgressions should be judged more modifiable with age.

However, in line with their suggestion that moral and conventional concepts are constructed because of different social interaction events associated with moral and conventional transgressions, Turiel, Smetana and Nucci have observed that the responses of children and adults to moral and conventional transgressions do differ. This literature will be reviewed in the following section.

1.6.1.2: Behavioral responses to transgressions

There have been nine studies that have compared the behavioral responses of children and adults to moral and conventional transgressions (Dunn & Munn, 1987; Edwards, 1987; Much & Shweder, 1978; Nucci & Nucci, 1982a; Nucci & Nucci, 1982b; Nucci & Turiel, 1978; Nucci *et al.*, 1983; Smetana, 1984; Smetana, 1989). Two of these field studies were conducted outside the Western cultural context (Edwards, 1987; Nucci *et al.*, 1983). All of these studies will be reviewed in this section apart from the Edwards (1987) study which was criticised in section 1.4.2.3.

Generally, the paradigm has involved observation of the children, after an initial familiarization period, for one or more periods of 45 minutes or longer. Children have been observed in school (Much & Shweder, 1978; Nucci & Nucci, 1982a; Nucci & Turiel, 1978; Nucci *et al.*, 1983; Smetana, 1984), home (Dunn & Munn, 1987; Smetana, 1989) and play area (Nucci & Nucci, 1982b) contexts. Transgressions were defined to have occurred if they evoked some form of response from a child or adult. The forms that this response might take are described in Table 1.7. The transgressions, themselves, were classified into two, or three, categories. All studies distinguished moral from conventional transgressions. Moral transgressions were defined as those transgressions which involved the justice, welfare or the rights of individuals or groups (e.g., a person intentionally hits another or a person takes the property of another). Conventional transgressions were defined as those transgressions which involved the disruption to social interaction and social order (e.g., engaging in work or play activities in other than the designated areas or time periods or engaging in acts different from the group while in group activities). In two of the studies (Much & Shweder, 1978; Nucci *et al.*, 1983), two types of conventions were identified; conventional school regulations (e.g., the conventional items mentioned above) and

general conventions (e.g., children engaging in opposite sex activities, table manners and clothing norms). The age of child observed ranged from 13 months to 14 years.

Table 1.7: Behavioral responses to transgressions.

Response category	Definition of response category
Injury or loss statement	Statements indicating pain or injury to self or personal loss (loss of property or personal space).
Emotional expressions	Statements expressing emotional state or exclamation of affect
Physical retaliation	Any physical act taken toward the transgressor
Rights statements	Statements by others regarding rights or fairness
Perspective taking requests	Statements by others pointing out the feeling of the victim; telling victims of transgression to express feelings to transgressor or telling transgressor how it feels to be the victim of the act.
Disorder statement	Indication that behaviour is creating a mess, disorder, or chaos.
Rule statement	Statements specifying a rule governing the action.
Sanction statement	Statement indicating that a sanction will be the response to the behaviour.
Command	A statement to do or cease from doing an act, without a statement of the rule.
Providing rationale	Reasons given for a rule or behaviour
Focus on care-giver's feelings	Statement to others pointing out the feelings of teachers regarding the transgressor or the act.
Physical restraint	Any physical act taken to restrain the transgressor or inhibit physical response.

Dunn and Munn (1984) followed a different methodology from that used in the other studies to be described. Their study was longitudinal (the child was observed at the ages of 18, 24 and 36 months), their classification of transgression types was different and they did not differentiate forms of maternal response. They did, however, find that mothers at all these ages were more likely to offer justifications for disputes over rights than over other topics (they did not, however, differentiate forms of maternal response). They also found that children at 18 months were more likely to show anger and distress in disputes over rights than over other topics (though this was not the case for the older children).

Smetana (1984; 1989) also observed children at this age (13 to 40 months) both in the home

(Smetana, 1988), like Dunn and Munn (1984), and in daycare centres (Smetana, 1984). At this age, children respond to transgressions predominantly as victims; they very rarely respond to conventional transgressions at all. Their responses to moral transgressions were the responses of victims and consisted of *injury/ loss statements*, *emotional expressions*, and *physical retaliation* (though the oldest subjects, at 37 months (Smetana, 1989) were beginning to make *rights statements*. Adults, whether they were care-givers (Smetana, 1984) or mothers (Smetana, 1989), differed in their responses to moral and conventional transgressions; to moral transgressions they made *rights statements* and *perspective taking requests*; to conventional transgressions they made *disorder statements*, *rule* and *sanction statements* and *commands*. Smetana (1984) also observed that care-givers, responding to conventional transgressions, were more likely to respond to conventional transgressions by *providing rationales* and by focusing on the *care-giver's feelings* for the younger children (aged 13 to 27 months) (and with *physical restraint* for the older children (aged 18 to 40 months).

Nucci (Nucci & Turiel, 1978; Nucci *et al.*, 1983[11]) and Much & Shweder (1978) observed slightly older children (2-10 to 5-2 years) in preschool contexts. Nucci and Turiel (1978) reported that children, at this age, did not respond to conventional transgressions. However, Nucci and Turiel did not distinguish between general conventions and school rules (see above). Nucci *et al.*, (1983) and Shweder and Much (1978) did. Both of these studies report child responses to conventional, as well as moral, rules. Children responded to moral transgressions at this age as they had done before; predominantly as the victims of the transgression with *injury/ loss statements*, *emotional expressions* and *physical retaliation* (Nucci and Turiel also note that the child at this age uses *commands* and they *involve adults*). Child responses to conventional transgressions consisted of *rule statements* (they also responded by *involving adults* and *commands* but these two responses were also used in response to moral transgressions). As with the younger children, adults responded by *perspective taking requests*, and *rights statements*. In contrast to the findings with younger children, adults were more likely to *provide rationales* in response to a moral transgression. Adults responded to conventional transgressions as they had in the studies on younger children with *disorder statements*, *rule statements*, *sanctions* and *commands*.

[11] It should be noted that the findings of Nucci *et al* (1983) are particularly interesting because they demonstrate that this differentiated reaction to moral and conventional transgressions is constant across cultures.

Nucci and Nucci (Nucci & Nucci, 1982a; 1982b) observed older children still (age range 6-10 to 14 years) in the classroom (Nucci & Nucci, 1982a) and in unsupervised play areas (Nucci & Nucci, 1982b). By this age the responses of children to transgression situations has changed; their responses are more similar to those of care-givers. Many child responses to moral transgressions were obviously still as victims (children responding with *injury/ loss statements* and *physical retaliation*; though not with *emotional expressions*). However, children were also making *perspective-taking responses* (though not in the playground encounters; see Nucci & Nucci, 1982b) and *rights statements* like adults though they did not *provide rationales* as the adults did. Children were responding to conventional transgressions like adults with *disorder* and *rule statements* and with *commands*. Children also responded to conventional transgressions with *ridicule*.

1.6.1.3: The cognitive model

In addition to their description of the social interactions and the resultant self-constructions that give rise to moral and conventional concepts, Turiel and his colleagues (Turiel, 1983a; Turiel & Davidson, 1986; Turiel & Smetana, 1984; Turiel *et al.*, 1987) have also developed a cognitive model of the mind they envisage accommodates these concepts. This "cognitive model" will be described below.

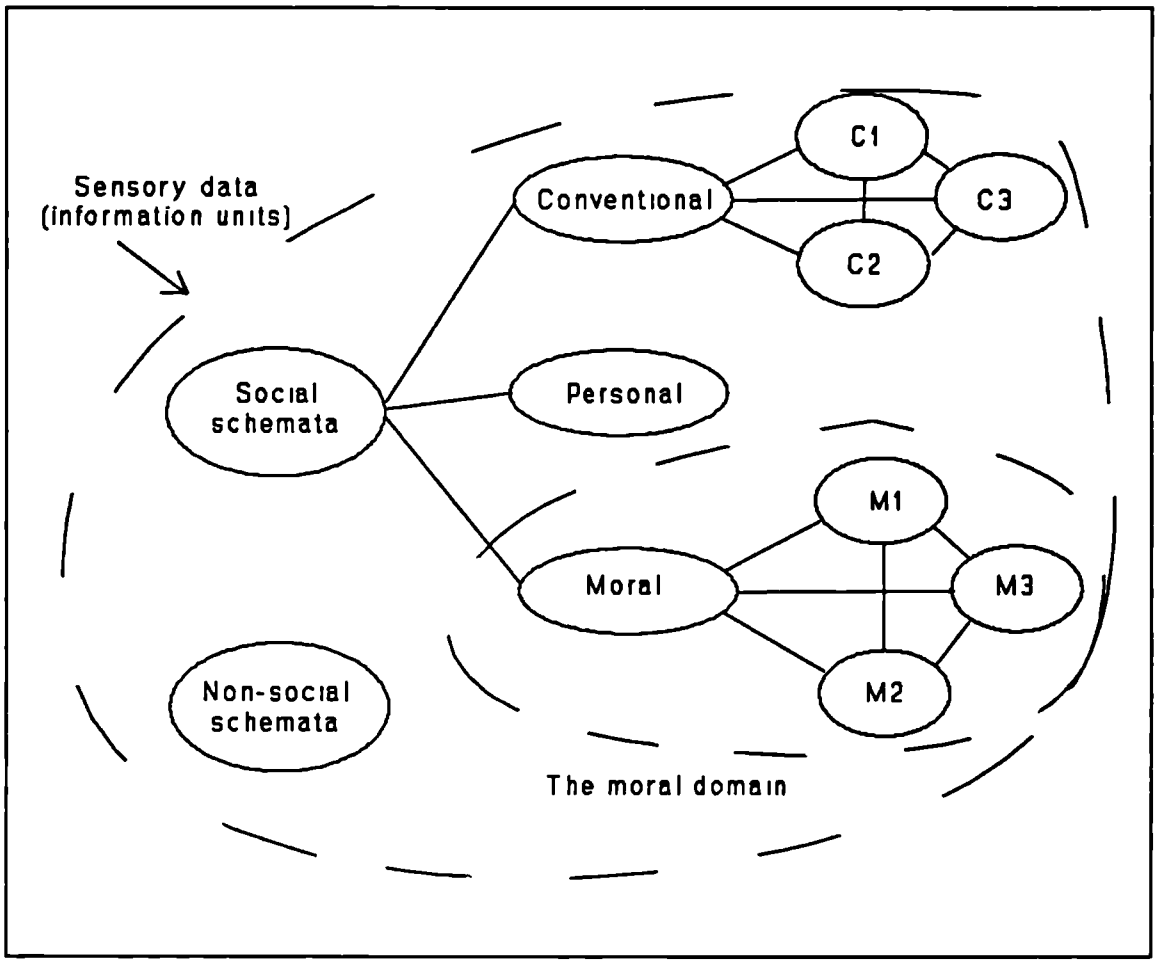
Turiel and Davidson (1986) claim that knowledge has three levels of structure in the mind: realms, domains and concepts. Units generated through subject-subject interaction enter the social realm, those through subject-object interaction enter the non-social realm. The realms are assumed to be subdivided into domains. The social realm is divided into the personal, conventional and moral domains. Units concerned with the empirical properties of people enter the personal domain. The personal domain involves all those prescriptions that are followed for advantage to the self, e.g. cleaning teeth twice a day. Units concerned with the forms of organisation with which the people are embedded enter the conventional domain. Units concerned with the intrinsic and obligatory forms of organisation of the subject-subject interactions themselves, leading to moral knowledge, enter the moral domain. Concepts, the third level of structure, are viewed by Turiel as the functioning substructures of the domains. Within a domain, they are supposed to be highly interdependent. These concepts can be seen as intuitive theories regarding the experienced social events (Turiel *et al.*, 1987).

Turiel and Davidson (1986) claim that the properties of a unit of information entering the mind determine which realm that unit of information is assimilated to. Assimilation is defined as the modification of elements to fit existing structures. Figure 1.1 is an attempted representation of Turiel and his colleagues' position within the schema framework of Mandler (e.g., Mandler, 1984). Within this framework, Mandler (1984) proposes a process of "meaning analysis" which refers to the "activation and accessibility of those schematic representations that best fit the available evidence" (Mandler, 1984; p. 126). Figure 1.1 represents realms, domains and concepts as clusters of interrelated schemata. An information unit, when being processed through meaning analysis would activate those schemata associated with it. An information unit detailing subject-subject interactions would activate the social schema. If the information unit concerned the well being of another, schemata making up the moral domain would be activated. Turiel states that concepts are the functioning substructures of the domains. In figure 1.1, this is assumed that this implies that concepts are those schemata "within the domains" (activated by the domain schemata) that can become consciously accessible. Their high "interdependence" is represented in figure 1.1 by the fact that they are all mutually excitatory.

Turiel *et al.*, (1987) argue that, given this emphasis on conceptual transformations and constructions, domains should not be seen as being unrelated to each other. Turiel argues that domains should be seen as analytically distinct structures that are related in multifaceted situations that include moral and nonmoral components (Turiel *et al.*, 1987). In figure 1.1, this is represented by the fact that the schemata that make up the individual domains do not excite schemata in other domains. Components of any given event such as the available information, the specific situational circumstances, and the assumptions about reality (e.g. natural order) will affect the ways the domains of judgement are accessed. Thus, Turiel does seem to be treating the domains as modular in the sense that the processing of one is independent of the processing of another. Their apparent relationship is due to the fact that the properties associated with a particular information unit may activate more than one domain; there are similarities of concepts stored in two, or more, domains.

Because domains are not unrelated to each other it is possible that a single event might excite more than one domain. Many events are not exclusively moral or conventional (Turiel, 1983a; Turiel & Davidson, 1986; Turiel & Smetana, 1984). A given information unit may activate more than one domain and more than one domain may provide inferences about a given information unit. Turiel (1983a) gives the various possible types of combinations:

Figure 1.1: A representation of Turiel and his colleagues' "cognitive model".



1) A straight forward mixture of domains (whether in conflict or in synchrony) requiring that both moral and conventional considerations be taken into account, e.g. when moral decisions are in contradiction to social organisational functions of coordination and efficiency (*see* Killen, 1985).

2) When issues are ambiguously multidimensional, so that significant discrepancies exist in their domain attribution by different people.

3) When the violation of another's social conventions is considered morally wrong because the violation of that convention causes psychological pain to those who have the convention.

4) When an act is committed that cannot cause physical, or psychological, harm to people who are alive, but can cause hypothetical harm to the dead or to spirits (Turiel *et al.*, 1987).Turiel and colleagues (Turiel, 1983a; Turiel & Davidson, 1986; Turiel *et al.*, 1987) conceive that the domains are fundamental categories organising thought. Turiel and Davidson (1986) assume that these domains are self-regulating. By self-regulating, they apparently mean that the activity of any given domain is independent of the processing occurring elsewhere in the mind. For example, Turiel and Davidson claim that a given domain actively initiates the assimilation of novel information units and actively accommodates or reinterprets prior information units in the light of new evidence. This assimilation and accommodation is not initiated by any system outside of the domain. Turiel and Davidson claim that the assimilation process itself selects the information units that are appropriate and excludes elements that are not appropriate for any given knowledge structure (realm, domain, or concept).

1.6.1.4: Criticisms

The main criticism of Turiel's cognitive model are that: (1) there is no evidence for any aspect of the model and; (2) the account is completely under-specified. While there is good evidence that individuals distinguish between moral and conventional rules there is no available evidence in support of Turiel's cognitive model. Indeed, it is difficult to imagine what evidence could be presented. It might be expected that individuals might exist who, following cerebral damage or through a developmental abnormality, might lack one of the domains. However, no such evidence has been presented. More important, however, is the second criticism. Turiel does not detail how this structure of realm, domain and concept develops. He claims that self constructive processes give rise to the moral/ conventional distinction but it seems difficult to imagine how these give rise to the cognitive hierarchy he envisages. For these reasons it seems that unless the model is extensively revised it must be abandoned.

1.6.2: Shweder's Social Communication Theory

Shweder, at least now, also views himself as a constructivist (Shweder, 1990). He, like Turiel and colleagues, has stressed the role of social interactions in the development of morality.

Shweder has proposed (Shweder et al., 1987; Shweder and Much, 1987) that moral beliefs have their origin in the "messages and meanings implicitly conveyed through talk, conversation, discourse, and customary practice" (Shweder and Much, 1987; P. 198). "Local guardians of the moral order" (parents, teachers, peers) convey to children "powerful morally relevant interpretations of events" (Shweder and Much, 1987). They are conveyed implicitly and, presumably, explicitly through the "verbal exchanges - commands, threats, sanction statements, accusations, explanations, justifications and excuses - necessary to maintain routine practices" (Shweder and Much, 1987; p. 203). The local guardians of the moral order trace the "boundaries of a normative reality and assist the children at stepping into the frame" (Shweder and Much, 1987; p. 203). Shweder and Much (1978) report example conversations of moral socialisation. From these it can be seen that norms were predominantly transmitted implicitly. To turn these into theories and boundaries it is necessary for the child to reconstruct the information. Shweder suggests "the inferences children draw about the moral (its form) and what's moral (its content) are personal reconstructions ..." (Shweder and Much, 1978; p. 198).

In contrast to Turiel, Shweder (Shweder and Much, 1987; p. 203) argues that "the emerging moral understandings of the child are the product of continuing participation in social practices (the mundane rituals of everyday life), and that those socially produced and reproduced understandings are the grounding for the child's later attempts reflectively or self-consciously to reconstruct his own moral code". Thus, according to Shweder, the initial concepts that are then used by the child in the reconstructive process to generate further concepts are supplied by cultural influences. None of the "basic building blocks" that the child uses to make sense of the world are provided by the reconstructive process itself.

In addition, Shweder argues, unlike Turiel, that this cultural process of specifying the "boundaries of normative reality" may result in there being more than one "natural moral law" (Shweder, 1990). Shweder (1990) has described three natural moral laws, or codes as he terms them, claiming that "a distinction needs to be made between moral arguments based on appeals to harm, rights, and justice (code 1) versus moral arguments based on appeals to duty, hierarchy, and interdependency (code 2) versus moral arguments based on appeals to natural order, sacred order, tradition, sin, and personal sanctity (code 3)" (Shweder, 1990; p. 2064). Shweder claims that the three codes are "three moral discourse realms, and they seem to pick up different aspects of the self or different ways of conceiving the self" (Shweder, 1990; p. 2064). Presumably, Shweder is arguing that the different "basic building blocks" supplied by different cultures may

result in different self constructions made by the child. These different self constructions presumably give rise to different "normative realities" and thus different "natural moral laws".

Obviously there are difficulties with Shweder's account. There is no evidence that people hold alternative "normative realities" for example. Certainly, there is no evidence, despite Shweder's efforts, of any culture that does not distinguish moral and conventional rules. The account is completely under-specified; it is difficult to imagine that it can make any predictions at all. In addition, there is a further problem that is also applicable to Turiel's account. This is the lack of evidence that children intellectually construct prescriptions for themselves in the way both Turiel and Shweder assume. Does emotion have no role in the development of moral prescriptions as Turiel *et al.*, (1987) claim? With this last question in mind, I move on to section 1.6.

1.7: Emotion theories; theories of the development of morality (2)

In section 1.5. I described three accounts (Piaget's, Youniss' and Kohlberg's) of the development of morality that could not predict the existence of the moral/ conventional distinction. In section 1.6, I described two accounts (Turiel and his colleagues' and Shweder's) that were designed to explain the moral/ conventional distinction. I mentioned the fact that both of these accounts ignore emotion. In section 1.7, I wish to describe one account of the moral/ conventional distinction that has stressed the role of emotion; Kagan's model.

1.7.1: Kagan's model

Kagan (1984) is interested in the generation of standards. He believes that the existence of standards is implied by the application of evaluative language to "thoughts and acts that can produce anxiety, shame, guilt or pride" (p. 112). He notes that what is positively, and negatively, evaluated in any given state, and at any given time, tends to change with the state and the time. However, he argues that "it is emotional states that form the bases for a limited number of universal moral categories that transcend time and locality" (Kagan, 1984; p.118-9). In other words, the evaluations may change but the way that these evaluations are generated, through emotional arousal, is part of the normal functioning of the developing mind. The emotional states he claims form the moral categories are:

1) Anticipation of the different types of anxiety in response to possible physical harm, social approval or task failure.

2) The feeling of empathy to those that are at risk.

3) The feeling of responsibility that follows the perception that one has caused harm to another.

4) The feeling of fatigue and/ or ennui that follows the repeated gratification of a desire.

5) The feeling of uncertainty that follows encounters with discrepant events that are not often understood, or reveal inconsistencies in thinking.

Kagan (1984) writes that

> "because people do not like to feel afraid, to feel sorry for someone less privileged, or to feel guilty, bored, fatigued, or confused, these unpleasant states will be classified as bad; and people will want to replace, suppress, or avoid them. The acts, motives and qualities that accomplish these goals will be good and, therefore, virtuous" (Kagan, 1984; p. 120).

The first sentence states that the unpleasant nature of certain states results in them being classified as bad; the labels of the states are all put within the category bad. The second sentence states that the actions that prevent the unpleasant states are classified as good; the labels of the actions are put within the category good. It would seem more sensible just to state, however, that the actions that lead to a bad/ good emotional state and the bad/ good emotional state itself are all classified as bad/ good. It should also be noted that the actual classification process, itself, is not explained. However, it seems that, according to Kagan, if the committing of an act leads to an unpleasant emotional state the act will be classified as bad, and if it leads to a pleasant emotional state it will be classified as good.

Given Kagan's position that it is the emotional reaction of the child to certain circumstances that results in the generation of standards, Kagan (1984) argues that emotion is the cause of the

moral/ conventional distinction. Moral transgressions are "tied to strong emotion" (Kagan, 1984; p. 121). Thus, Kagan argues, when the emotional reaction to the standard vanishes, the standard becomes conventional. Promiscuity is no longer thought, in Western society, to be immoral because it is not tied to strong emotion. There is some evidence for this position. Arsenio and Ford (1985) observed that children (aged 6 - 8 years) attributed a significantly more negative emotional state to agents who committed the transgression and to themselves witnessing the transgression if the transgression was moral rather than if it were conventional. They also demonstrated that inducing a negative affective state aided the recall of stories concerning moral, but not conventional, transgressions. An induced negative or positive state did not result in significant differences in the number of the two types of transgression recalled.

However, it does not really seem that the moral/ conventional distinction can simply be conceptualised as a distinction between two groups of transgressions of which only one is "tied to strong emotion". Dressing up in opposite sex clothes has been found by Turiel (1983a) to be treated as a conventional transgression yet to be observed in opposite sex clothing would result in a strong emotional state of embarrassment. It is quite possible to believe that an individual could state that fidelity was a matter of convention, yet still be ashamed if their infidelities were known. Thus, Kagan's account does not seem to be an adequate account of the moral/ conventional distinction either.

1.8: Conclusions

In this first chapter I have described the nature of the moral/ conventional distinction and detailed several of the more dominant accounts of the development of morality. The following conclusions can be made from this:

1) The moral/ conventional distinction exists; subjects reliably distinguish between moral and conventional transgressions in their judgements.

2) Criticisms of the distinction's existence, whether they are theoretically or empirically based, are invalid.

3) The early theories of the development of morality (Piaget's, Youniss' and Kohlberg's) could not predict the existence of the moral/ conventional

distinction and must be discarded.

4) The two constructivist accounts (Turiel and his colleagues' and Shweder's) of the moral/ conventional distinction suffer from chronic under-specificity and a lack of clear predictions.

5) Kagan's model of the moral/ conventional distinction cannot account for the existence of conventional transgressions which are "tied to strong emotion".

In summary, it seems that while the moral/ conventional distinction exists, accounts of its existence are highly unsatisfactory. Chapter two will present a new model of the development of the moral/ conventional distinction.

Chapter Two

A Model of the Development of Morality

2.1: Introduction

In chapter one, I reviewed the literature on the moral/ conventional distinction. This literature clearly reveals that children, from the end of the third year, judge moral transgressions differently from conventional transgressions. In addition, I described six theories of the development of morality. I described three formulations which would not predict the existence of the moral/ conventional distinction (e.g., Piaget, 1932; Youniss, 1980; Kohlberg, 1969). I described two formulations which, though designed to explain the moral/ conventional distinction, make few testable claims about its development (e.g., Turiel, 1983; Shweder, 1990). I described one formulation which attempted an account of the distinction based on emotion which must be flawed because of its claim that no conventional transgressions are tied to strong emotion (Kagan, 1984).

In this chapter, I intend to develop a model of the development of morality. This model will be expressed as a Developmental Contingency Model (DCM; *see* Morton, 1986) and as an information processing model. I thus intend to make clear claims about: the prerequisites for the development of various aspects of moral behaviour (the inhibition of violent action; the moral emotions and; the moral/ conventional distinction) and; the cognitive architecture that underpins these aspects of moral behaviour. The information processing model will be grounded in the Supervisory Attentional System Framework (SASF). This framework will be described in section 2.2. An understanding of emotion is necessary for a theory of the development of morality. For this reason, I will present two models of emotion in section 2.3; Mandler's (e.g., 1984) account and Oatley and Johnson-Laird's (1987) account. Section 2.4 will consist of a description of the cognitive architecture that is involved in the performance of the moral behaviours mentioned above. Section 2.4.1 will detail a processing model within the SASF. Section 2.4.2 will place a model of emotion, which borrows heavily from Mandler and Oatley and Johnson-Laird, within this processing model. Section 2.5 will detail the mechanism, the Violence Inhibiting Mechanism (VIM), that I claim is a prerequisite for all the moral behaviours listed above. In this section, I will explain the functioning of the mechanism, its developmental consequences and how its operation can be over-ruled.

2.2: The Supervisory Attentional System Framework (SASF)

The Supervisory Attentional System Framework (SASF) is represented in figure 2.1. This figure represents a framework for the description of the mental control of action. For some time a distinction has been made between controlled and automatic action (e.g. James's dichotomy of "ideo-motor" and "willed" acts; Shiffrin & Schneider's (1977) separation of controlled and automatic processing of perceptual inputs). Examples of types of automatic action are: the execution of a task without awareness of the performance (e.g. walking on a stretch of familiar flat ground); an action elicited without deliberate attention or awareness (e.g. drinking from a glass while in conversation); the absence of control over the drawing of attention to a particular object (e.g. the orientating response); an action that can be performed without interfering with other tasks. Shallice (Norman & Shallice, 1980; 1986; Shallice, 1982; 1988a; 1988b; Shallice & Burgess, 1991; *in press*) argues that the operation of two complementary processes (see figure 2.1) corresponds to the distinction between automatic and controlled actions; "contention scheduling" (CS) allows action sequences to run off automatically, without conscious control or additional resources; the "Supervisory Attentional System" (SAS) is involved in controlled actions.

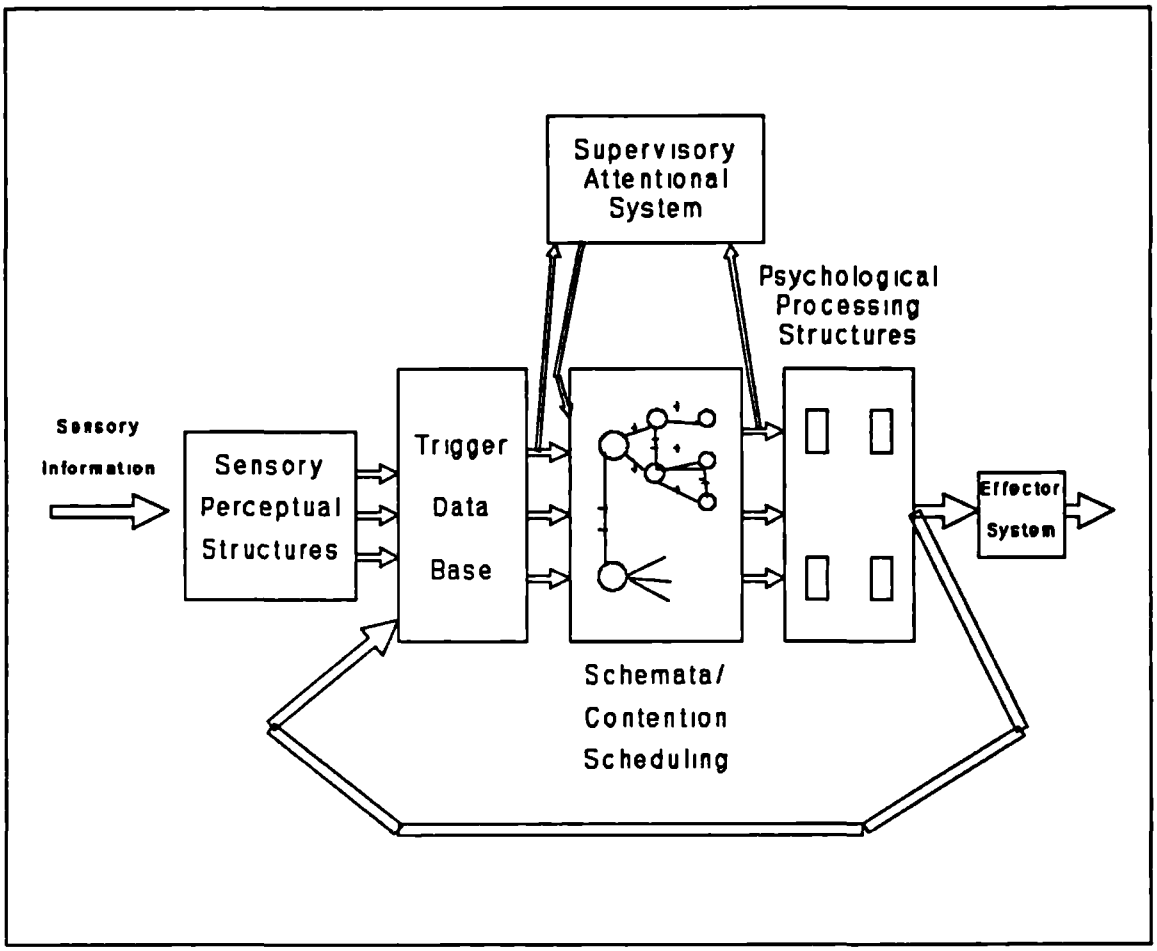
2.2.1: Contingency Scheduling

Shallice (1988a) proposes two types of schemata in contingency scheduling (see figure 2.1); source and component schemata. Shallice postulates that source schemata (e.g. for driving a car) call into operation particular component schemata (e.g. for braking the car). He also argues that a component schema for a particular source schema may be a source schema for other component schemata (e.g. the braking component schema will be a source schema for the component schemata of pressing the brake pedal and changing down the gears).

Shallice (Norman & Shallice, 1980; 1986; Shallice, 1988a) argues that a schema is activated if its activation level exceeds a given threshold. He suggests (Shallice, 1988a) that each schema has certain trigger inputs and that the presence/ absence of these trigger inputs is one of the factors which determines the activation level of the schema (see figure 2.1). Other factors will be discussed below. Shallice (1988a) proposes that the schemata are independently activated by triggers and are in mutually inhibitory competition for selection. Which schemata any particular schema inhibits is dependent on the particular processing systems the schema requires (Shallice,

1988a).

Figure 2.1: The Supervisory Attentional System Framework. According to Shallice (1988a), the figure "illustrates the control processes operating on the isolable subsystems and processing regions, rather than their on-line operation" (Shallice, 1988a; p. 334). Shallice claims that the *psychological processing structures* represent the bulk of the on-line operation of such systems. In *Schemata/ Contention Scheduling*, source schemata are represented by the apex nodes of each tree structure, and component schemata, by lower nodes. Excitatory links are labelled by a + and inhibitory links, by two parallel lines.



Thus, according to Shallice, the activation level of a schema is dependent on: (1) the concordance of environmental conditions with the trigger inputs; (2) the activation levels of schema which have excitatory (because they are source schema for the particular schemata) or inhibitory (because the schema depend on the same processing systems) connections and; (3) the operation of the SAS.

According to Norman & Shallice (1980; 1986) contention scheduling involves two basic processes: (1) the competition of source schemata with each other in the determination of their

activation value. This competition occurs through the excitatory and inhibitory links with each other. And; (2) the selection of a schema according to whether its activation value has exceeded the threshold. Thus, contention scheduling allows the selection of an action in response to environmental stimuli if the environmental stimuli sufficiently correspond to a set of target inputs such that the activation level of a schema will exceed the threshold of that schema.

2.2.2: The Supervisory Attentional System

Shallice argues that the SAS provides a second source of control on the operation of schemata and that it operates through the provision of excitation and inhibition to the activation levels of the schemata that are needed to be controlled to achieve a particular goal. Shallice proposes that when the operation of the SAS on a schema ceases, the activation level of that schema will decay back to the value that the other sources of activating input would produce.

It should be noted that this is all the framework states about the operation of the SAS; that it is a system that can activate schemata independently of external stimuli. Shallice (e.g., 1988a) has claimed that the SAS is also involved in the formation of plans but the sub-components of the SAS that might be involved in this plan formation have not yet been specified.[1] Thus, the SAS should be understood as a collective term for a variety of psychological functions all of which have been left unspecified. What has been specified is how these collective psychological functions interface with the schemata of contingency scheduling.

The actual operation of the SAS is also under-specified. Norman & Shallice (1986) give the types of information the SAS would require to operate:

> "representations of the past and present state of the environment, of goals and intentions, and of the repertoire of the higher-level schemata it could activate .... The system would need to know aspects of the operation of a selected schema or, to be more precise, of those selected schemata which it could potentially activate (source schemata). It would need to know not only which source schemata had been selected but also the action sequences they produced and probably the eliciting triggers as well. Without such information, error correction would be a hopeless task, but it is a key function of the supervisory system (Shallice & Norman, 1986; p. 14)".

---

[1] Attempts have been made to detail sub-components of the SAS (Shallice & Burgess, 1991; *in press*). However, these attempts have so far described plan formation as a single sub-component of the SAS.

How the various types of input (e.g. the representations of past and present state of the environment) needed by the SAS for its functioning are combined and utilised has not yet been considered.

However, for the purposes of this thesis it is not necessary for there to be a clear model of reasoning (i.e. SAS functioning) placed within the SASF. All that is necessary is a clear account of the control of action. This the SASF provides. A clear account of the control of action is necessary because of the claim that will be made later (section 2.5.1) that the fundamental root of all moral representations is an innately specified mechanism for the control of violent action. A clear account of the control of action is also necessary because of the importance of emotion in the control of action (*see* Oatley & Johnson-Laird, 1987). Other accounts of the control of aggression have implicated emotional reactions (particularly empathy) in its inhibition. This argument will not be followed here but it will be proposed later (chapter four) that emotional reactions are crucial in the development of moral meta-knowledge. Two models of emotion are detailed below.

2.3: Models of emotion

In section 2.2, I described the components of the SASF. In this section, I will detail two models of emotion: Mandler's (e.g., 1984; 1987; 1991) and; Oatley and Johnson-Laird's (Oatley, 1988; Oatley & Johnson-Laird, 1987). These two models were chosen because of their adaptability to the needs of the proposed account of the development of morality. Other approaches, such of those of Frijda (1986) and Izard (1971; 1977), that are not so adaptable to the goals of this thesis, will not be considered.

2.3.1: Mandler's model

Various theorists have proposed that emotion is a consequence of the interruption of behaviour (Dewey, 1895; Meyer, 1956; Paulhan, 1930; *see* Mandler (1984) for a historical review of these "conflict" theories). Mandler (1964; 1975; 1984; 1987; 1991) follows this tradition.

Mandler's position, his "discrepancy/ evaluation theory", proposes two underlying processes; autonomic (sympathetic) nervous system arousal and evaluative cognitions. He believes that

arousal is frequently a consequence of the occurrence of discrepancies in perception, action and thought. He states that "within the purview of schema theory, these discrepancies occur when the expectations generated by some schema (whether determining thought or action) are violated" (Mandler, 1991; p. 61). He proposes (Mandler, 1984) that situations which result in emotional reactions do so through the interruption of integrated or organised response sequences (e.g. the integrated response sequences of driving will be interrupted by the brakes not working). He proposes (Mandler, 1991) that there is a "difference detector" which is activated by these discrepancies between expectancies and occurrences in perception, action and thought. The activation of this "difference detector" results in autonomic nervous system (ANS) arousal. This ANS arousal is assumed to be undifferentiated experientially whatever the details of the discrepancy situation. In other words, the ANS arousal response would be experienced similarly whether the discrepant event is worse or better than the expected one; the ANS arousal responses that are "evaluated into" and provide intensity for joy or disappointment are not differentiated.

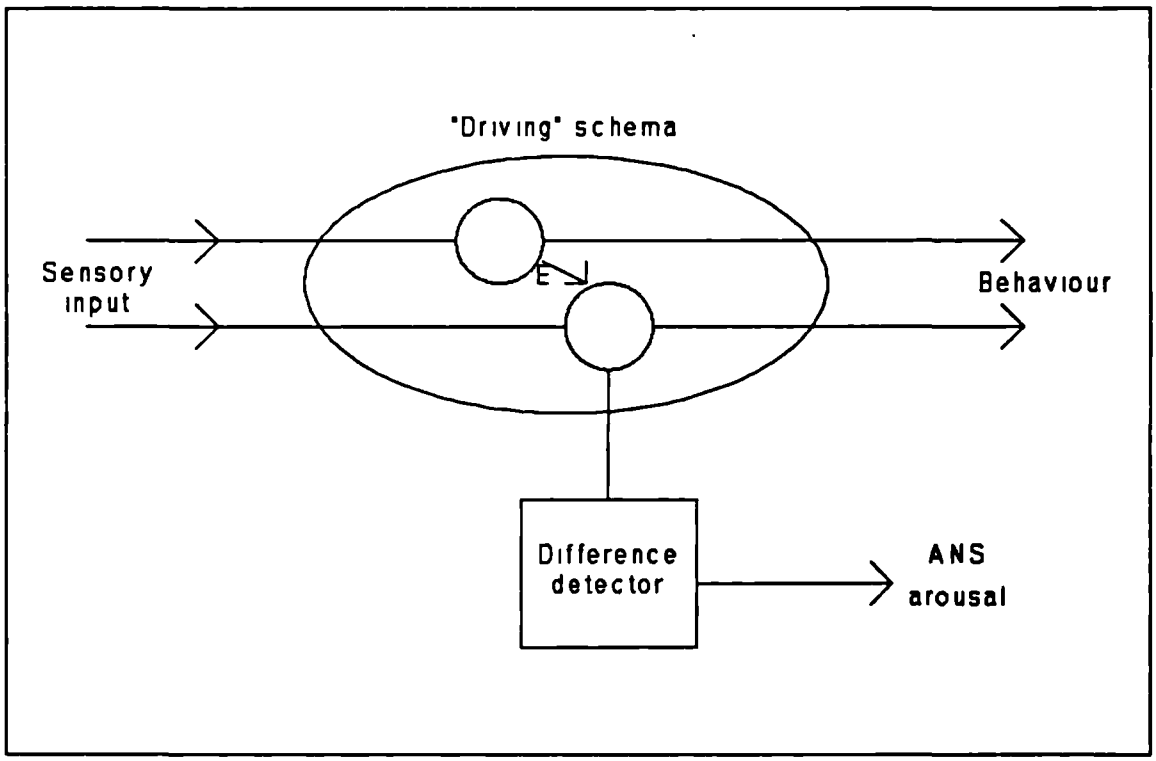
Figure 2.2 represents a very simplified version of Mandler's model; here the "difference detector" is shown operating only on the "driving" schema. It is assumed that sensory input entering the schema activates the first element of the schema. This element initially sends out commands which result in action; in this case, driving. This element also sends out a message to the second element detailing the current situation (i.e., the distance between the individual's car and the one in front). It is this second element which continues to control behaviour. However, this second element only continues to activate behaviour if the sensory input coming into the schema does not conflict with the expectations set up by the first element in the schema. If there is conflict between the inputs from these two information sources, the element activates the difference detector which results in the activation of the ANS and an undifferentiated arousal response.

There is now considerable empirical evidence that discrepancies do result in ANS arousal (e.g., Frankenhaeusur, 1971; 1975; Fry & Ogston, 1971; Mandler, 1991; Nakamura (cited in Mandler 1991); Sher, 1971). For example, Nakamura (cited in Mandler, 1991), presented subjects with very simple stories and asked subjects how they would end. The stories were presented one sentence at a time and the last sentence was either consistent or discrepant with the preceding context. Nakamura found significant increases in heart rate following discrepancy.

While it appears clear that discrepancies do result in ANS arousal, it is less clear that different

domains of emotion inducing stimuli do not result in differentiable arousal states. Ekman, Levenson and Friesen (1983) have provided evidence that there is emotion-specific activity in the ANS. They demonstrated that the production of facial expressions associated with particular emotions resulted in ANS activity specific to some of these emotions. However, the importance of this finding is unclear. It would seem quite possible that these differences in ANS activity are not accessible to the individual's awareness. The ANS response may be differentiable at the physiological level but not at the psychological level. Certainly, it seems that after some time period this arousal is unspecified; arousal engendered by one type of emotion stimulus (e.g. fear) at time 1 can affect the intensity of other affects that are a consequence of emotion stimuli (e.g.

Figure 2.2: A representation of Mandler's model of the generation of arousal. Sensory input entering the schema activates the first element of the schema. This element sends out commands which result in behaviour. This element also sends out an "Expectations" (E) message to the second element detailing the current situation (i.e., the distance between the individual's car and the one in front). It is this second element which continues to control behaviour. However, this second element only continues to activate behaviour if the sensory input coming into the schema does not conflict with the expectations set up by the first element in the schema. If there is conflict between the inputs from these two information sources, the element activates the difference detector which results in the activation of the ANS and an undifferentiated arousal response.

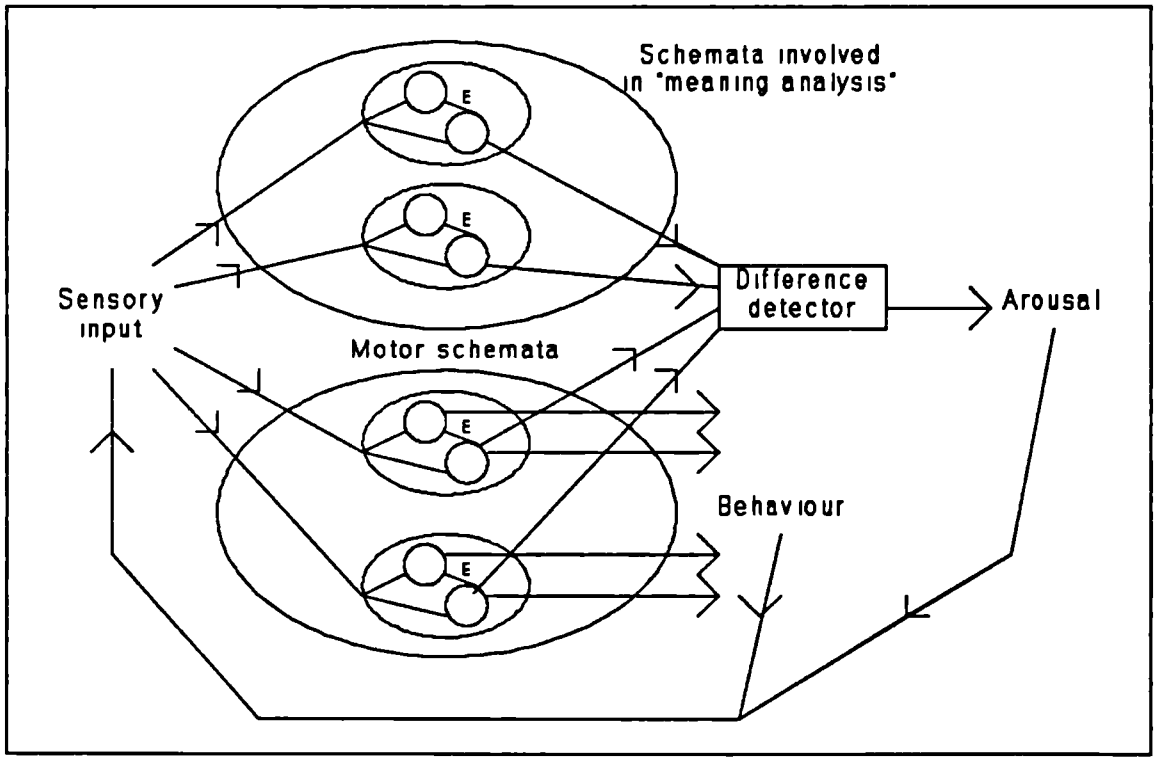


sexual arousal) presented at time 2 (Dutton & Aron, 1974; see Manstead & Wagner (1981) and

Reisenzein (1983) for reviews of excitation transfer).

Mandler claims that "the emotional content - the quality of the emotion that follows - is set by specific cognitive circumstances of the interruption and possibly its consequences" (Mandler, 1984; p. 46). He considers that the "existence of ANS arousal, together with some evaluative state, will produce a conscious emotional state in keeping with the current intentional state of the individual" (Mandler, 1984; p. 149). Mandler (1984) proposes a process of "meaning analysis" which refers to the "activation and accessibility of those schematic representations that best fit the available evidence" (Mandler, 1984; p. 126). This meaning analysis is seen to be engendered either by the arousal or the general situation and current cognitive states. In any given situation, interruption induced arousal will be coupled with the schemata activated through meaning analysis to produce a particular emotional experience (see figure 2.3).

Figure 2.3: A representation of Mandler's model of the experience of emotion. In figure 2.3, I have represented two schema stores: one of consciously accessible schemata, schemata which are actually involved in "meaning analysis" and; one of (probably mostly motor) schemata which are not involved in "meaning analysis". Both forms of schemata would be connected to the "difference detector".



Mandler (1984; 1991) also proposes that meaning analysis supplies the individual with values for objects and activities. Mandler (1984, 1991) claims that the individual has three sources of

value available to him:

(1) Innate approach and withdrawal tendencies. Mandler suggests that it is the secondary effects of these tendencies, the observation of our own approaches and withdrawals, that is one of the conditions that produces judgements of positive and negative value.

(2) Through cultural, social and idiosyncratic learning. Past experiences, not necessarily with the object (e.g., someone telling you frogs legs are revolting), can result in idiosyncratic values.

(3) Structural. Mandler (1991) suggests that this aspect of value resides in the cognitive structure of objects, in the relations among features, as in the appreciation of an object seen as beautiful or abhorrent as a function of a particular structural concatenation. Mandler (1991) claims that structural value arises out of our experiences with objects and the analyses of their constituent features. Mandler (1991) suggests that we find positive value in familiar objects.

2.3.2: Oatley and Johnson-Laird's model

Mandler's model focuses on the experience of emotion. In contrast, Oatley and Johnson-Laird's model focuses on the role of emotion in the control of action. Oatley and Johnson-Laird (1987; also, Oatley, 1988) believe that emotions have a central role in the management of multiple goals in an uncertain world. They define goals as "symbolic representations of possible states of the environment that a system will try to achieve" (Oatley & Johnson-Laird, 1987; p. 30). In a situation where there are multiple goals there will be a need for interruptions, when one goal, unexpectedly, becomes more urgent than the one that was currently controlling behaviour. They argue that these interruptions are responsible for emotions.

Oatley and Johnson-Laird (Johnson-Laird, 1983a; 1983b; Oatley, 1988; Oatley & Johnson-Laird, 1987) advocate a modular mind; i.e., the mind is made up of "quasi-autonomous processors, each of which is able to achieve a particular goal given certain inputs and preconditions" (Oatley, 1988; p. 351). These modules are seen to be hierarchically organised (Oatley, 1988a; 1988b; Oatley & Johnson-Laird, 1987); see figure 2.4. Only the conclusions of the top-level module, which contains a model of the whole system (i.e. a model of the self), becomes conscious (see

also Johnson-Laird, 1983a; 1983b). Each module is assumed to be monitoring the status of its goal (Oatley, 1988). When the likelihood of the success of this goal changes an emotion may be induced. Oatley and Johnson-Laird (1987) suggest five status changes to goals that will induce emotional reactions: sub-goals being achieved, failure of major plan or loss of active goal, active plan frustrated, self-preservation goal threatened, and the gustatory goal being violated.

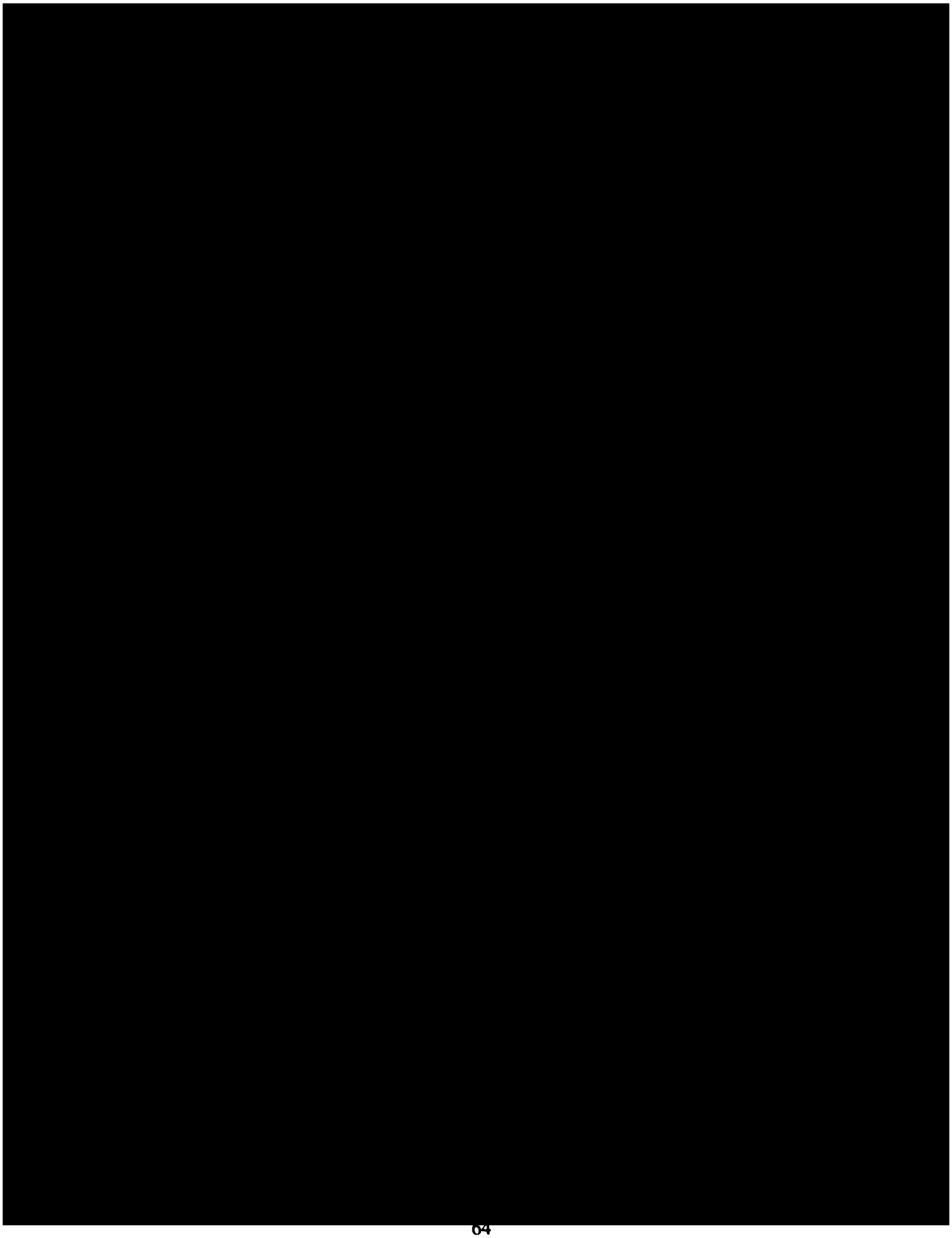
Within Oatley and Johnson-Laird's (Johnson-Laird, 1983b; Oatley, 1988; Oatley & Johnson-Laird, 1987) modular mind, communication occurs in two ways:

1) Propositional. Propositional signals are symbolic: i.e. "they have internal structure that plays a part in denotation within the system" (Oatley & Johnson-Laird, 1987; p. 32). These signals are seen to pass from one module to another, thereby invoking the recipient to some form of action (see figure 2.4).

2) Non-Propositional. Non-propositional signals have no internal symbolic structure. They do not denote anything. They simply function causally. They propagate globally among the processors to set them into specific modes.

Oatley and Johnson-Laird (1987) propose that emotions are based on non-propositional communications; additionally, they propose that the main function of emotions is communicative (Oatley, 1988a; Oatley & Johnson-Laird, 1987). Each module is assumed to be monitoring the status of its goal (Oatley, 1988). When the likelihood of the success of this goal changes, a non-propositional signal will be sent out from this module which interrupts the on-going plan, and changes the current goal priorities. This non-propositional signal puts the system into a mode of functioning (see figure 2.5), hence Oatley and Johnson-Laird (1987) use the term emotion modes. "For an emotion to occur, the system needs to be in one emotion mode or oscillating between two. The intensity of an emotion corresponds to the amount of the system entrained in a particular mode and to the subsequent degree of locking into that mode" (Oatley & Johnson-Laird, 1987; p. 34).

Following the work of Ekman (e.g. Ekman, Friesen & Ellsworth, 1982), Oatley and Johnson-Laird (1987) propose that there are five emotion modes corresponding to the five basic emotions Ekman argues for. However, Oatley and Johnson-Laird claim that the full experience and expression of an emotion (i.e. the distinctive phenomenological tone, the somatic changes, the



behavioral expressions, and the courses of action) is more than just the emotion mode. An emotion mode is only a necessary, but not sufficient condition, for emotional experience. "In adults, the full emotion typically also includes a conscious evaluation of the juncture in planning, based on propositional signals reaching the operating system so that it is able to ascribe a meaning to the emotion mode, and so that the voluntary action can be scheduled" (Oatley & Johnson-Laird, 1987; p. 34). The emotional experience consists of propositional and non-propositional communication within the system.

In addition to the five basic emotions (happiness, sadness, fear, anger and disgust) that correspond to the five emotion modes (when the appropriate propositional signals have been sent to the top-level, conscious module), Oatley and Johnson-Laird (1987) propose that there are complex emotions. These complex emotions are complex in the sense that they are founded on a basic non-propositional, emotion mode, but have a propositional evaluation which is social and includes reference to the model of the self. They argue that these complex emotions are dependent on an appraisal in which performance is compared with that which is expected on the basis of the model of the self. Oatley and Johnson-Laird give the example of remorse where, they claim, "an aspect of the sense of self is lost: the model of self suffers a decrement in its attributed positive characteristics" Oatley & Johnson-Laird, 1987; p. 47). An implication of the above account is that the range of possible interpretations of an emotional state is narrower than Mandler (1984) considered. The range of possible interpretations of an emotional state would only be those congruent with the basic emotion mode.

2.3.3: Contrasting Mandler with Oatley & Johnson-Laird

There are six points where Mandler and Johnson-Laird may be contrasted:

Firstly, and most obviously, the theories are based in very different frameworks. Mandler's "discrepancy/ evaluation theory" is located within his schema theory framework (*see*, Mandler, 1984). Oatley and Johnson-Laird's model is located within Johnson-Laird's modular mind framework (*see*, Johnson-Laird, 1983a). While some of the lower level modules in Johnson-Laird's framework seem to function similarly to schemata (i.e., they are activated by sensory input, conduct some form of computational process and then have eventual behavioral implications), not all do. In particular Johnson-Laird's topmost module, the one which contains the model of the self, cannot be considered to be a single schema.

Secondly, the mechanisms implicated in the activation of emotional reactions in each of the models are very different. Mandler's model considers that an arousal response is a consequence of a discrepancy in perception, action or thought being "noticed" by a "difference detector". Oatley and Johnson-Laird consider that an emotional state is engendered by a change in the status of a goal that is being monitored by one of the modules in the mind.

Thirdly, there is disagreement over the nature of the arousal response. Mandler claims that the arousal response is undifferentiated; that, whatever the discrepancy conditions, the arousal response alone does not determine the conscious experience of the emotion. Oatley and Johnson-Laird claim that the arousal response is appreciably different according to the nature of the goal status change that has occurred. Oatley and Johnson-Laird claim that there are five emotion modes that are experienced appreciably differently[2] (as happiness, sadness, fear, anger and disgust).

Fourthly, while both theories postulate an evaluative process, the specifics of this process are rather different. Mandler proposes that this evaluative process consists of a process of meaning analysis which refers to the activation and accessibility of those schematic representations that best fit the available evidence. Oatley and Johnson-Laird propose that the evaluative process consists of the transmission of a propositional signal from the affected module to the module responsible for consciousness. For complex emotions (e.g., remorse) the propositional evaluation is social and includes reference to the model of the self.

Fifthly, the two theories differ in the stress they lay on behaviour. Since, Mandler is more interested in investigating the experience he refers to behaviour only implicitly. Oatley and Johnson-Laird stress behaviour as their interest is that of demonstrating the role of emotion in coordinating action in an uncertain world.

Sixthly, the two theories differ in their considerations of the function of emotion. For Mandler,

---

[2] Oatley and Johnson-Laird (1987) claim that for an emotion to be experienced it is necessary that an emotion mode (of which the arousal is an aspect) is activated and a propositional signal is sent from the affected module to the module responsible for consciousness. This, of course, implies that the five emotion modes are experienced appreciably differently. However, when Oatley (1988) considers the consequences of an emotion mode not being sent, he states that there will be a feeling of dread without knowing the cause of this dread; i.e., the arousal response is not experienced appreciably differently but for the full understanding of an emotional experience it is necessary that a propositional signal is sent from the affected module.

the arousal induced by behavioral interruptions alerts the system to the interruption and expedites the system's remedial action. For Oatley and Johnson-Laird, the function of the emotion modes is to aid fast response to the interruption by immediately switching the system into a particular pre-programmed mode "designed" to deal with that form of interruption.

2.4: A model of emotion within the SASF

Given the purpose of this thesis, a model of the development of morality, I have no intention of becoming embroiled in a debate on the relative merits of the two positions on emotion described above. Mandler's position is disadvantaged by its inability to explain all forms of emotional experience; particularly happiness under goal completion situations. Oatley and Johnson-Laird's position in disadvantaged by their reliance on the Ekman data indicating different physiological patterns for different emotions. Both positions are disadvantaged by their under-specification. Furthermore, both positions are disadvantaged, for the purposes of this thesis, by being placed within rival frameworks of the mind to the one being followed here; the SASF. For these reasons I will now describe a model of emotion that is located within the SASF. This model should not be seen as a rival to the two positions described above. If anything this model should be viewed as a reformulation of Mandler's position (with reference to Oatley & Johnson-Laird's position) within the SASF. There are differences between the model to be described here and Mandler's position but these may reflect nothing more than the structural differences between the two frameworks the two models are located in. However, before I can describe a model of emotion within the SASF, it is necessary to develop a processing model within the SASF which includes those elements necessary for an account of emotion. I will attempt to do this in the following section. Please note that this initial model is not designed for the generation of predictions. Its purpose is as an explanatory tool. It will allow the expression, within an information processing model, of concepts that are necessary for an account of emotion.

2.4.1: A model within the SASF

Norman and Shallice (1986) define attention phenomenologically as the output from the SAS, i.e. what you are attending to is the item that is producing the output from the SAS. However, this says nothing about what attention actually is. However, two recent models have specified the attentional process (Duncan & Humphreys, 1989; Houghton & Tipper, 1992a; 1992b). They

divide the attentional process into three component stages: the formation of the perceptual description; the selection process and; the result of the selection process.

During the formation of the perceptual description a segmented array of structural units is generated by the relevant sensory system. During the selection process, the array of representations is matched against a set of target specifications. Structural units having properties which fail to match the target specification are inhibited. Structural units having properties that match the target specification are excited. The most excited structural unit is selected. According to Duncan and Humphreys, the selected structural unit is then passed into visual short term memory (VSTM). Duncan and Humphreys claim that only structural units in VSTM can become the focus of behaviour. Houghton and Tipper propose that the selection process allows "binding" to take place. They define "binding" as the setting of the variable parameters that are necessary for the initiation of action schemata (for example, it is necessary to set an expected weight if an object is to be lifted).

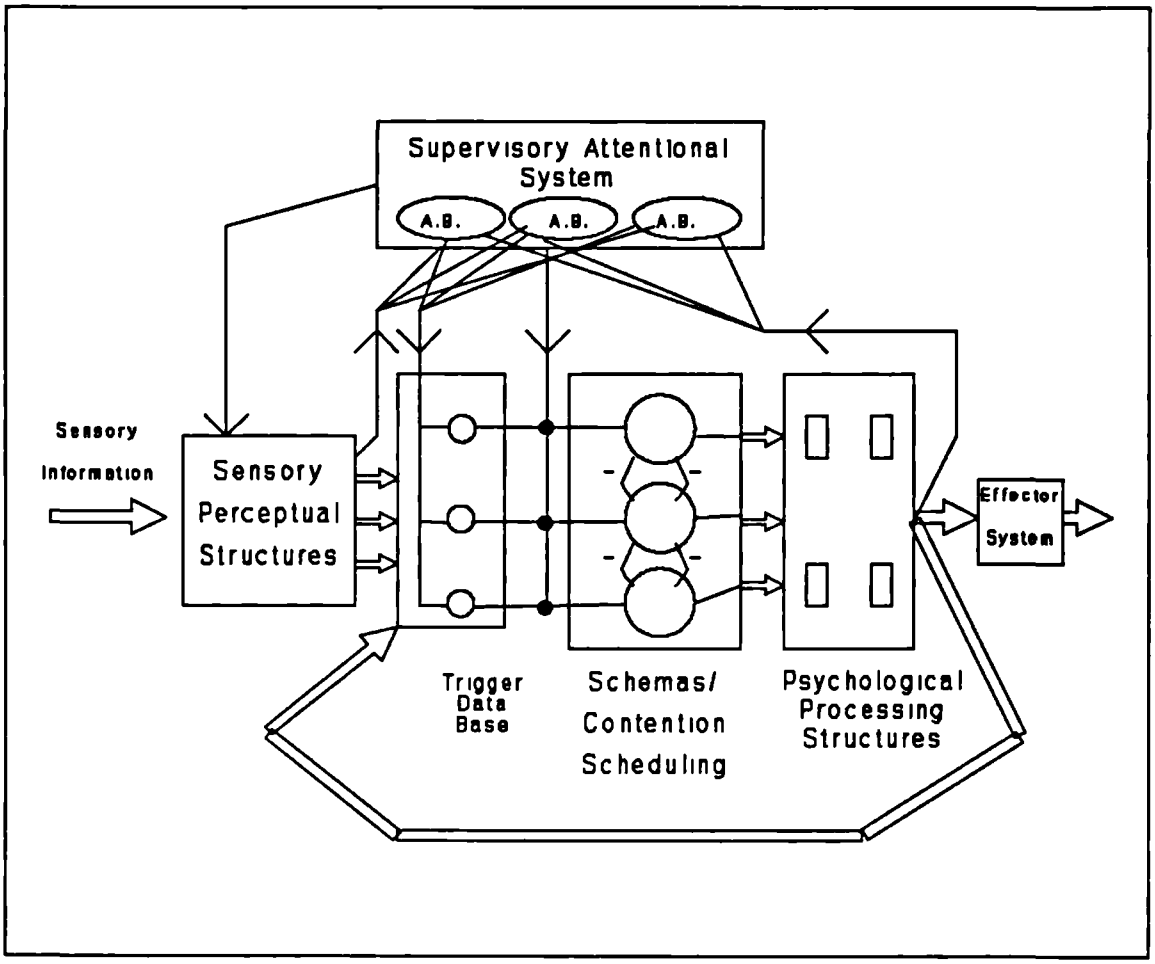
If the SAS is involved in the control of behaviour, the SAS must be controlling the attentional process; it must be setting the target specifications detailing which structural unit is to be attended to. If the SAS was not setting the target specifications it could only control which act was to performed by activating the relevant schema. The object the act would be performed upon would be the object that was most salient to the trigger data base. Figure 2.6 shows the SAS specifying the target specifications to the sensory perceptual structures. Duncan and Humphreys claim that the output of the visual perceptual process is the selected structural unit's entrance into the VSTM. In figure 2.6, it is assumed that the outputs of the perceptual processes for all of the sensory structures are passed into "awareness buffers" contained in the SAS.[3] It is the structural units in the awareness buffers that the functions of the SAS operate on; i.e, the structural units may be "decoupled" by a Theory of Mind Mechanism, manipulated (e.g, as in the mental rotation tasks) or changed (as in counterfactual reasoning). It is assumed that it is these "awareness buffers" that are damaged in blindsight (*see* Weiskrantz, 1986) or in cases of neglect (e.g., Bisiach, Capitani, Luzzatti & Perani, 1981).

Work with blindsight and neglect patients indicates that while they cannot use visual

[3] Please note that I do not wish to claim that "conscious awareness" is the sum of these "awareness buffers". However, I would assume that conscious experience reflects some product of these awareness buffers.

representations for planned activity, visual representations are processed for certain automatic

Figure 2.6: A model within the SASF. In this model the SAS interfaces with the rest of the system in three ways: (1) by sending target specifications to the *sensory perceptual structures*; (2) by transmitting representations from the *awareness buffers* to the *trigger data base* for *binding* and; (3) by activating schemata in *contention scheduling*. *Contention scheduling* interfaces with the SAS by transmitting representations to the awareness buffers of the SAS.



actions. If shown two houses in their neglected area, one of which is on fire, they will not be able to identify the objects but they will be able to say which one they would prefer to live in. While the preferred choice may be either of the two houses (the fire may be either aversive or rewarding), the individual subject's responding is consistent; they always prefer the burning house or the normal house (Marshall & Halligan, 1988). Taking Mandler's (1984) position on value, the fire may initiate a withdrawal response (as burning) and, through "meaning analysis", a negative value for that house or the fire may initiate an approach response (as heat source) and, through "meaning analysis", a positive value for that house. The response that is activated in a particular individual will presumably be determined by their developmental history. It is

assumed that these withdrawal and approach responses are automatic. The representation initiating the response would not be passed through the awareness buffers in the SAS. The representation would be being passed straight from the sensory perceptual systems to the trigger data base. Thus, attended representations are passed to the "awareness buffers" while the perceptual array is passed onto the trigger data base. Thus, structural units that have not been attended to may still activate behaviour by being in the perceptual array that is passed onto the trigger data base.

Meaning analysis within the present model involves the "binding" of representations to each other (*cf.* with notions of "Central Coherence"; Frith, 1989). Taking the above example of the neglect patients, it is assumed that these patients were operating on a single information unit; "fire". However, it is assumed that normals were operating on a representation formed through "meaning analysis"; the representation of the fire and the representation of the house had been "bound" together to form a representation of "burning house". Of course, Mandler's notion of "meaning analysis" is much more complex than this. To conceptualise Mandler's ideas on meaning analysis it is necessary to have a model of consciousness. However, a model of consciousness is not needed for this thesis.

To control behaviour, the SAS must activate the relevant schemata. In addition, the SAS must be transmitting a representation of the attended representation to the trigger data base to allow "binding" to occur. These representations, following meaning analysis, will contain more information (variable parameters for "binding") than the information units passed from the sensory perceptual structures to the trigger data base. This transmission of representations from the SAS to the trigger data base is represented in figure 2.6. It is also assumed in figure 2.6 that Contention Scheduling may result in the transmission of representations to the awareness buffers in the SAS, for example during recall. Taking the Headed Records framework for memory retrieval (e.g., Morton, Hammersley & Bekerian, 1985), the Description would be the structural units in the awareness buffer at the time of the attempted recall. The Description would be "bound" against the Headings of the trigger data base and a recalled Record would be transmitted from the CS to the relevant awareness buffer for its representation.

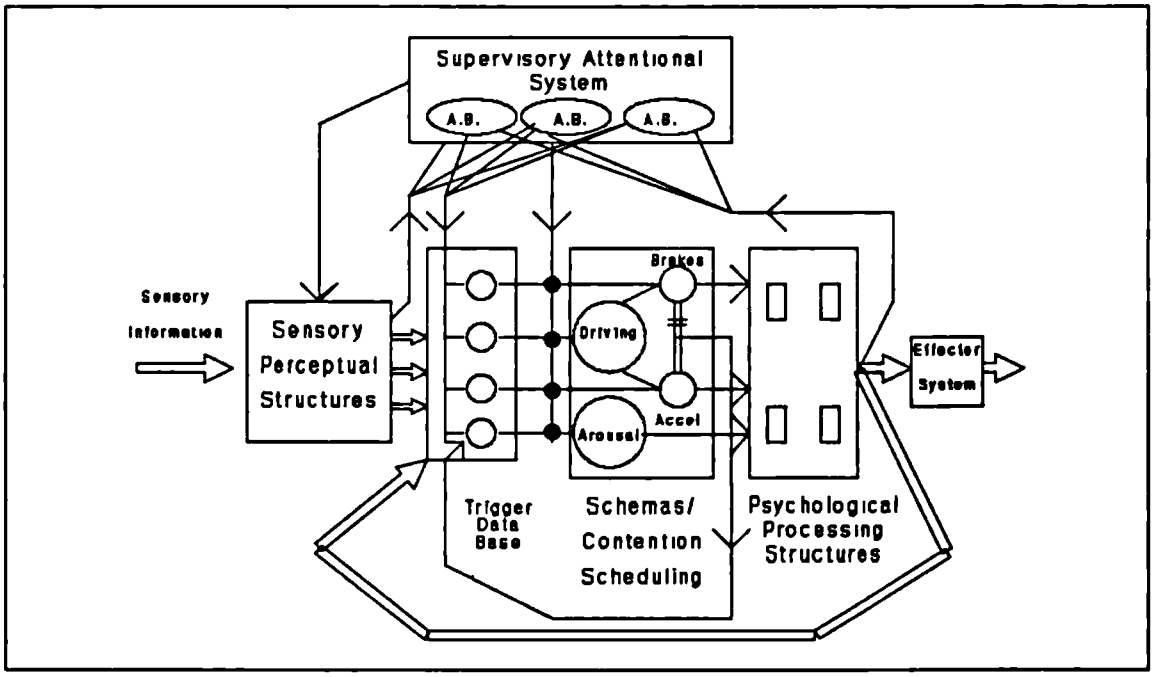
2.4.2: A model of emotion

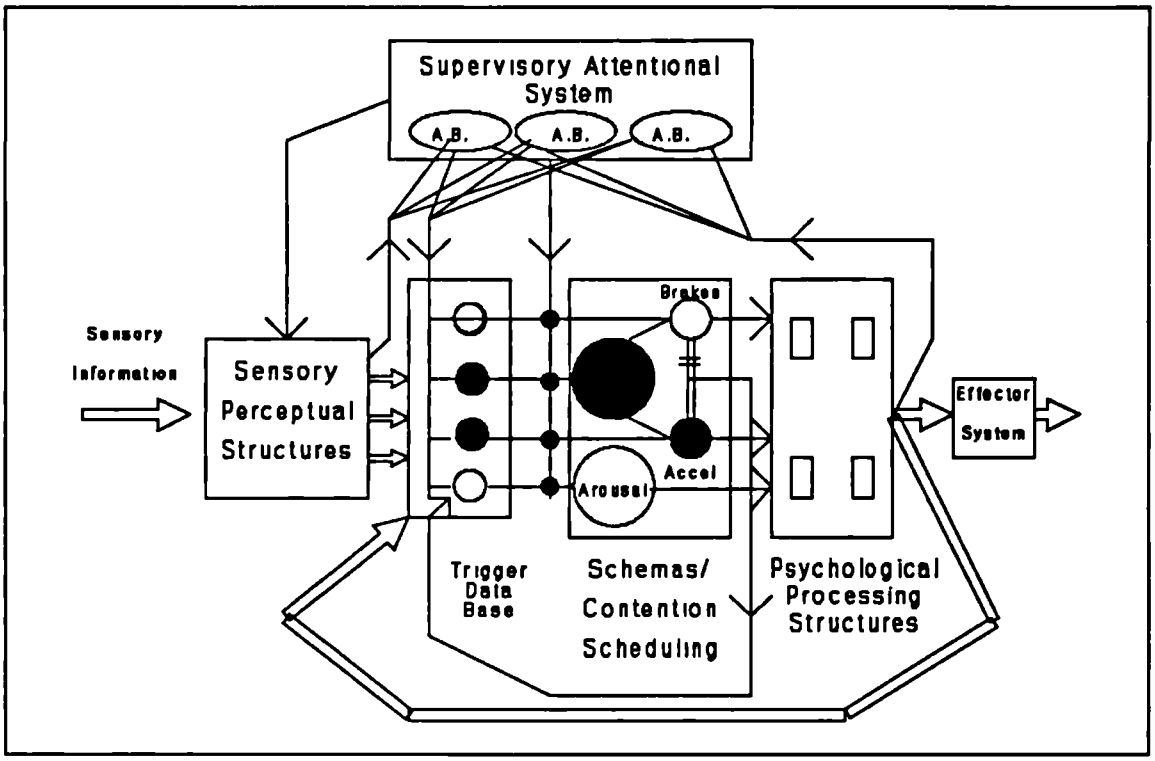
The model to be presented here is another "conflict" theory of emotion; this model assumes, like

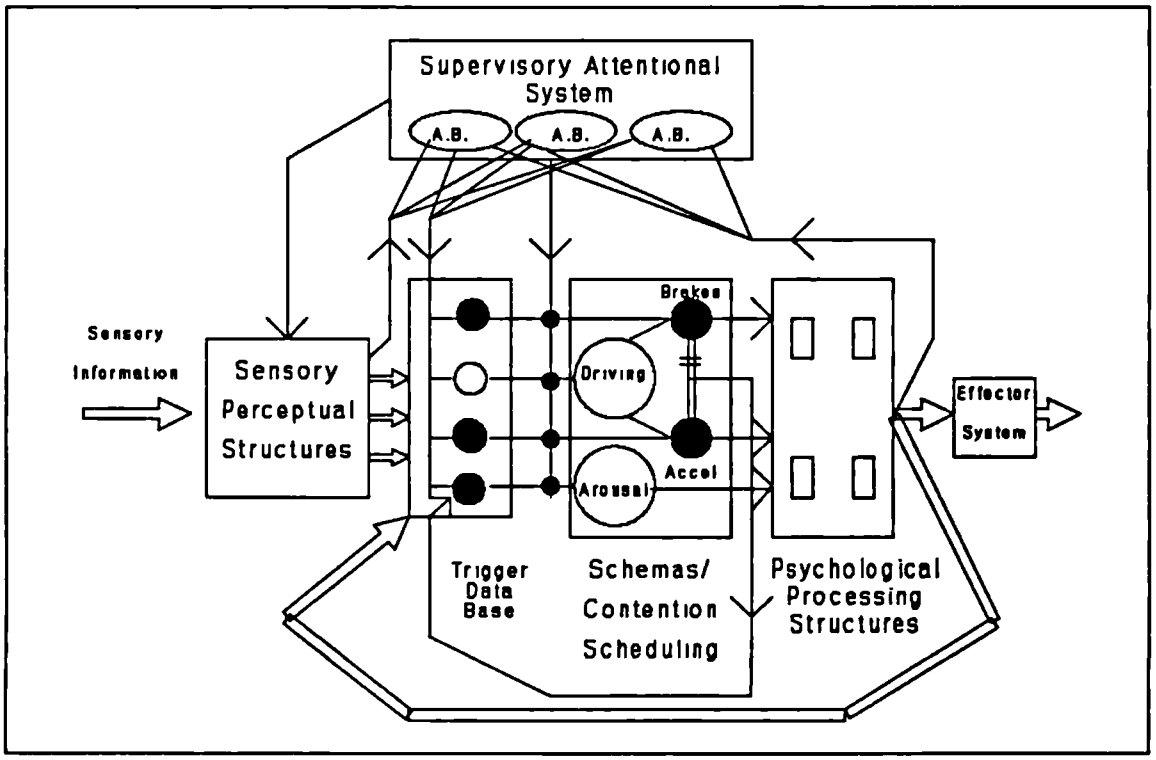
those described in section 2.3, that emotional arousal is induced by interruptions to behaviour. Within the SASF, behaviour, as expressed by the effector systems, is the product of an activated schema. Which schema is currently active in contingency scheduling is determined either by the stimuli passing through the sensory systems to the trigger data base or direct SAS involvement and also by a competitive process of excitation and inhibition between the schemata (see section 2.11). I suggest, like Mandler, that there is a "difference detector", which when activated, excites an arousal response. I suggest that this "difference detector" is sensitive to the inter-schema competition process.

The operation of this detector is detailed in figure 2.7. Figure 2.7(a) depicts an extended version of the general processing model presented in section 2.41. Here, the individual is assumed to possess four schemata. One schema, the driving schema, is represented as a source schema for the two component schemata responsible for braking and accelerating. The connections between this source schema and the component schema allow it to inhibit and excite these component schema. It is assumed that the two component schemata are mutually inhibitory. The fourth schema, when activated, is assumed to result in ANS arousal.

Figure 2.7: A model of emotion. 2.7(a) shows the system at rest. 2.7(b) shows the system when the conditions necessary to activate the driving schema are present. 2.7(c) shows the system under conditions of inter-schema competition. The "arousal schema" has been activated.




Supervisory Attentional System
A.B.
A.B.
A.B.
Sensory Information
Sensory Perceptual Structures
Brakes
Accel
Arousal
Effector System
Trigger Data Base
Schemas/ Contention Scheduling
Psychological Processing Structures


Supervisory Attentional System
A.B.
A.B.
A.B.
Sensory Information
Sensory Perceptual Structures
Brakes
Driving
Accel
Arousal
Effector System
Trigger Data Base
Schemas/ Contention Scheduling
Psychological Processing Structures

At time one (represented by figure 2.7(b), it is assumed that the driving source schema is currently active. This schema activates and inhibits the component schemata of braking and accelerating according to the information that is being passed to the trigger data base. If the road is clear, the accelerator schema may be activated. This is the situation represented in figure 2.7(b).

Under normal circumstances, if the situation changes (e.g., there are traffic lights on red some way ahead) the driving schema will begin to inhibit the accelerator schema and excite the brakes schema. However, the braking schema will **not** be activated by its trigger node since the lights are still some distance away. The driving schema will be performing a planning function (though, I stress that it will not be forming a plan). After some time, and at the appropriate distance to the lights, the activation level of the braking schema will cross its threshold (either as a result of the operation of the driving schema or the trigger node of the braking schema being activated) and the individual will begin to brake.

Figure 2.7(c) represents the state of the system under emergency conditions. It is assumed that the road has been clear and the driving schema has been activating the accelerator schema. However, a car turns out of a nearby side street. At this moment the node for the braking schema is triggered, activating the braking schema. However, the accelerator schema is still active (its activation level has not yet decayed sufficiently). Thus, at this moment, the two component schema of the driving schema are maximally, mutually inhibitory. It is assumed that this maximal, mutual inhibition is sufficient to increase the activation level of the arousal schema across its threshold thus allowing this schema to become active. An undifferentiated arousal response will be activated.

It is assumed here, following Mandler, that how an arousal response is experienced will be determined by the process of "meaning analysis" described in section 2.4.1. Other representations (i.e., of the situation) will be "bound" into the representation of the arousal response. The resultant, more complex representation will be the basis of the experienced emotion.[4] Also, following Mandler, it is assumed that the principle functions of the arousal response are: Firstly, to alert the individual (or, more importantly, his SAS) to behavioral

[4] I say the basis for the experienced emotion given my earlier claim that consciousness must involve some calculation on the representations in the awareness buffers.

interruption and; Secondly, to ready the system for action.

2.5: A model of the development of morality

It should be clear from chapter one that any complete theory of the development of morality must have considered the following four issues:

(1) The development of the moral/ conventional distinction.

(2) The changes in moral meta-knowledge over time.

(3) The details of the cognitive structures that underpin the observed development of morality.

(4) The existence of the various emotional responses to moral and conventional transgressions.

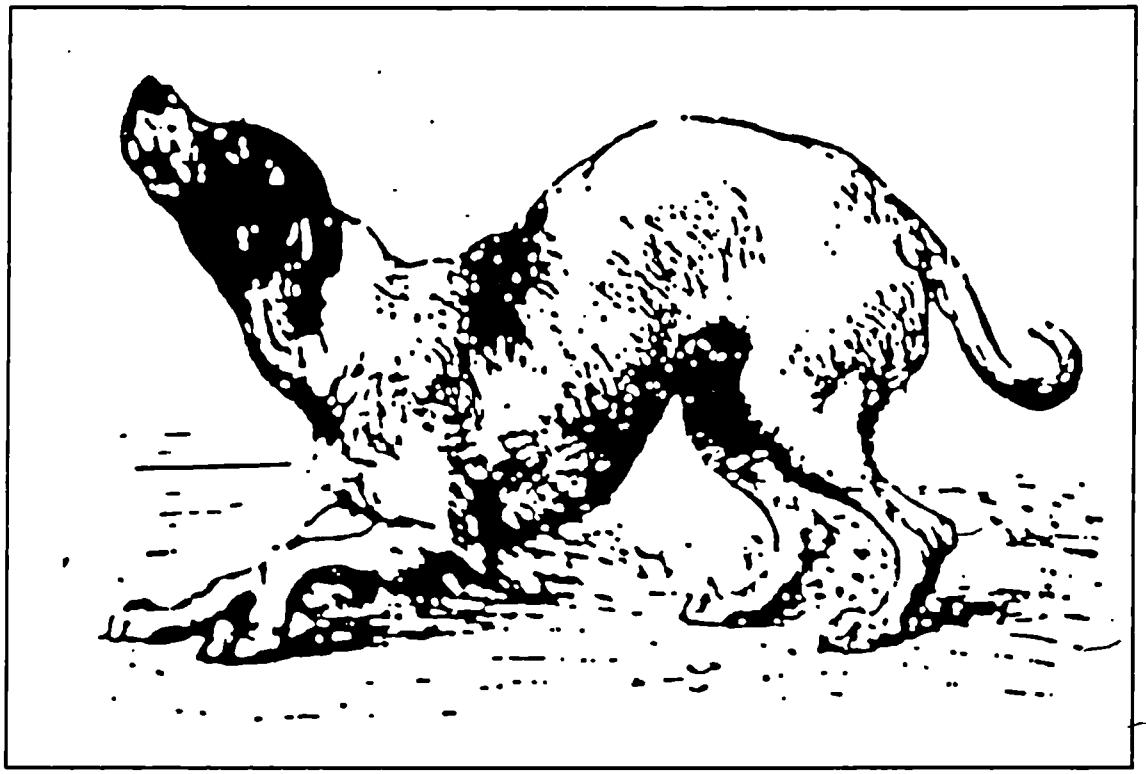
This thesis does not purport to describe a complete model of the development of morality. However, I do intend to detail: an account of the development of the moral/ conventional distinction; the cognitive structures that underpin, and are prerequisites for, the development in morality and; the emotional responses associated with moral transgressions. I have no intention of supplying an account of the changes in meta-knowledge over time. The influence of some of the cognitive structures to be described on moral meta-knowledge will be mentioned in chapter four but this will be in an on-line, as opposed to developmental, context. Also, I am only interested in the specifically moral emotions (i.e., guilt, remorse, conscience pangs, sympathy and emotional empathy). I will not consider those emotions associated with both moral and conventional transgressions (i.e., shame, embarrassment).

2.5.1: The Violence Inhibiting Mechanism (VIM)

Several ethologists have suggested that animals have developed mechanisms to control their aggression (Eibl-Eibesfeldt, 1970; Lorenz, 1966; 1981). They note that the display of submission cues to a conspecific aggressor frequently results in the termination of the attack. For example, dogs when attacked by a stronger opponent bare their throats (figure 2.8). This

results in the cessation of the fight. These ethologists have claimed that humans lack these mechanisms. I claim that they are wrong. I claim that there is an innately specified schema which inhibits all behaviour in the context of non-verbal communications of distress (i.e., sad facial expression, the sight and sound of tears); *see* Fridlund (1991a; 1991b) for more general work on the non-verbal communicatory role of the emotion facial expressions. I term this schema, the Violence Inhibiting Mechanism (VIM).

Figure 2.8: Submission cue exhibited by a dog.



While there is no evidence that humans terminate their attacks following the victim's display of distress cues, there is evidence that observers become aversively aroused when witnessing the distress of others.[5] Several studies have demonstrated that perceived distress in others generates an aversive emotional reaction that can be measured as physiological arousal in observers (e.g., Bandura & Rosenthal, 1966; Berger, 1962; Craig & Lowery, 1969; Krebs, 1975). Also, this aversive reaction to distress cues seems to be present from an early age; infants cry to the sound of crying. Simner (1971) reported this reactive crying in 2- to 3- day old children, a finding that

[5] It should be noted that Mandler's emotion model must have difficulty accounting for evidence that subjects are aroused by observing others in distress. It seems extremely difficult to explain this arousal in terms of the violations of expectancies. What expectancies could be being violated?

has been replicated by Sagi & Hoffman (1976). This reactive crying is not simply a response to noxious auditory stimuli; infants do not cry to equally loud and intense non-human sounds (Simner, 1971).

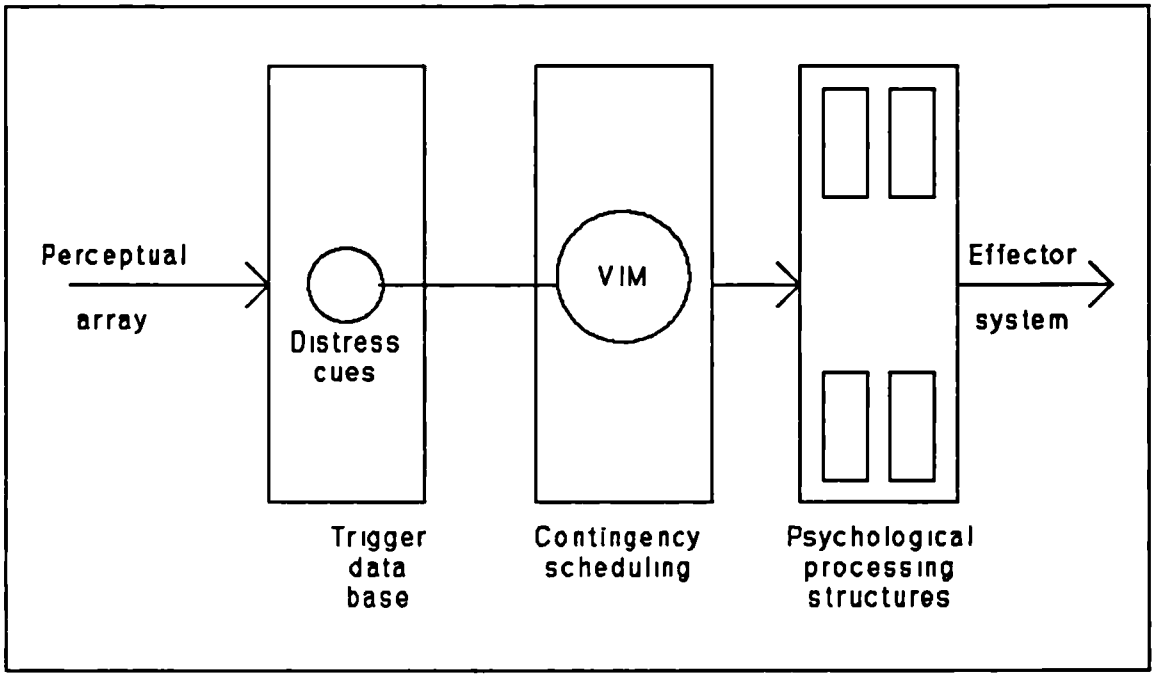
Obviously, I do not claim, given the model of emotion presented in section 2.4.2, that VIM directly induces this arousal. Instead, I claim that VIM, when it is activated by a distress cue representation being transmitted from the sensory perceptual mechanisms to the trigger data base, inhibits all other schemata (i.e., inhibiting all behaviour). Excitation of the inhibitory connections between the schemata activates the trigger for the schema responsible for autonomic arousal. An unspecified arousal response, which may be evaluated into sympathy, is produced. This arousal response is experienced aversively because of the withdrawal tendencies associated with the inhibition of behaviour (following Mandler's (1984) claims about the formation of value).

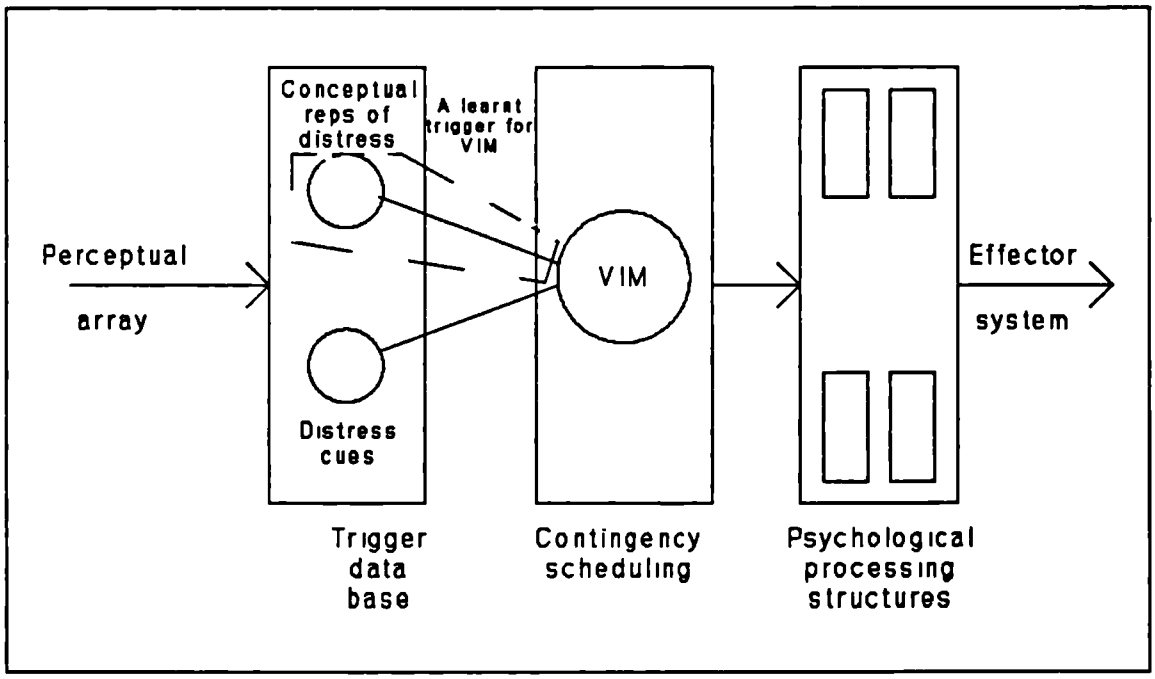
I also suggest that VIM develops functionally over time. This development is represented in figure 2.9. Initially, the functioning of VIM is innately specified; the inhibitory schema is activated following the stimulus of distress cues. However, I suggest that a process of classical conditioning results in the generalization of VIM activation. As said above, initially VIM is activated by the presence of distress cues; distress cues are the unconditioned stimulus (US), VIM activation the unconditioned response (UR). However, the US of distress cues will be frequently paired with various conceptual representation CSs; e.g., the presence of a victim, conceptual representations of the victim's sadness, representations of the transgression. Frequent pairings of the US of distress cues with the CS of victim conceptual representations will result in the CS activating a CR; activation of VIM. Additional triggers for VIM will have been learnt.

2.5.2: VIM and Empathy

In section 2.5.1, I made reference to several studies that have demonstrated that perceived distress in others generates an aversive emotional reaction that can be measured as physiological arousal in observers (e.g., Bandura & Rosenthal, 1966; Berger, 1962; Craig & Lowery, 1969; Krebs, 1975). I used these studies to infer the existence of VIM. However, these studies have instead been used previously as evidence for empathy in observers (e.g., Batson, 1988; Batson, Fultz & Schoenrade, 1987). In this section, I will explain the relationship of empathy to the operation of VIM.

Figure 2.9: Showing the development of VIM. Figure 2.9a shows the innately specified VIM. Representations of distress cues activate the trigger, thus activating VIM. Figure 2.9b, shows the development of VIM following the pairing of conceptual representations of distress with distress cue representations; effectively, VIM can be activated by an additional trigger - by conceptual representations of distress. New triggers for VIM have been learnt.





Empathy has been defined as "an affective response more appropriate to someone else's situation than to one's own" (Hoffman, 1987; p. 48). It is most frequently considered to be generated as a consequence of perspective (or role) taking. As Batson states "Perspective taking is the psychological variable most often assumed to be the antecedent of specifically empathic reactions to another's distress" (Batson *et al.*, 1987; p. 172). Like VIM, empathy has also been implicated in the control of violent action (e.g., Feshbach, 1983; Perry & Perry, 1974; Samenov, 1984; Yochleson & Samenov, 1986). These authors have suggested that empathy may prevent, or lead to the termination of, violent attacks. Indeed, Gibbs (1987) has suggested that emphasis should be given to the development of empathic reactions in violent offenders.

Empathy has also been implicated in the development of morality. As I stated in chapter one, Kagan (1984) has claimed that empathy is one of the "strong emotions" that are activated by moral, but not conventional transgressions, and which result in the moral/ conventional distinction. Hoffman (e.g., 1975; 1977; 1984; 1987) generated a description of a developmental progression in the way that individuals appreciate the existence of others and an analysis of how this developmental progression might influence the child's response to empathic arousal. However, since the development of empathy is not of primary relevance to this thesis, and since the data Hoffman presents is open to other accounts, this work will not be described. In addition, Hoffman (1984; 1987) has stressed the role of empathy as a motivation for action. Hoffman claims that people help in order to reduce this empathic reaction. Hoffman (1987) has also claimed that empathy has a role in determining which moral principles an individual will hold:

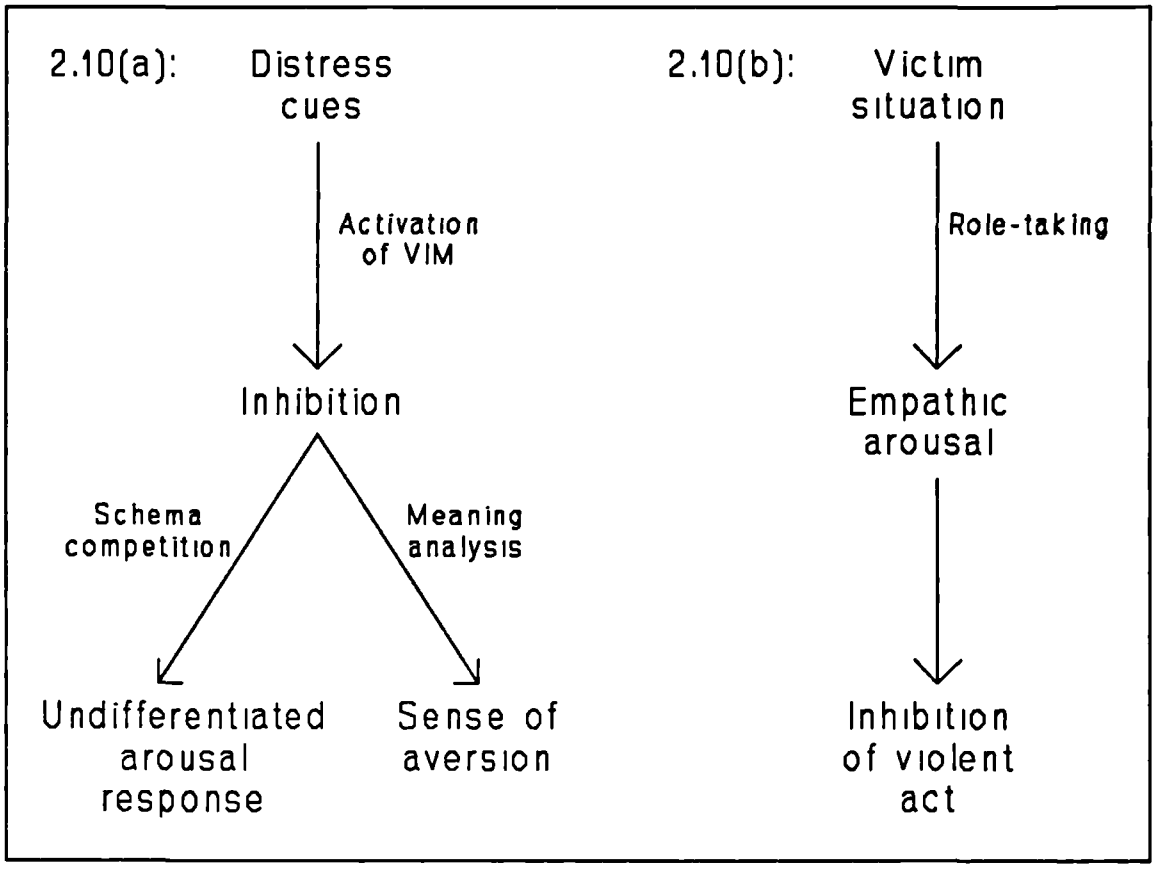
> "At some point in late childhood or adolescence the individual is exposed to various moral principles, usually in a loose, scattered fashion. The "cafeteria" model seems appropriate here: The more empathic one is, the more receptive one should be to caring, need, equality, and perhaps effort" (Hoffman, 1987; p. 63).

However, the "cafeteria" model will not be followed in this thesis.

Figures 2.10(a) and 2.10(b) contrast the operation of VIM and the empathic process in information flow diagrams. Figure 2.10(a) shows my claim that VIM is activated by distress cue representations. The operation of VIM results in the inhibition of action. This inhibition, through the schema competition process, excites arousal. In addition, this inhibition, as effectively a withdrawal response, is interpreted through meaning analysis as aversive. Figure

2.10(b) shows a representation of the empathic process. Victim situations activate the role taking process and this, in turn, results in empathic arousal. This empathic arousal is then assumed to result in the inhibition of the violent act.

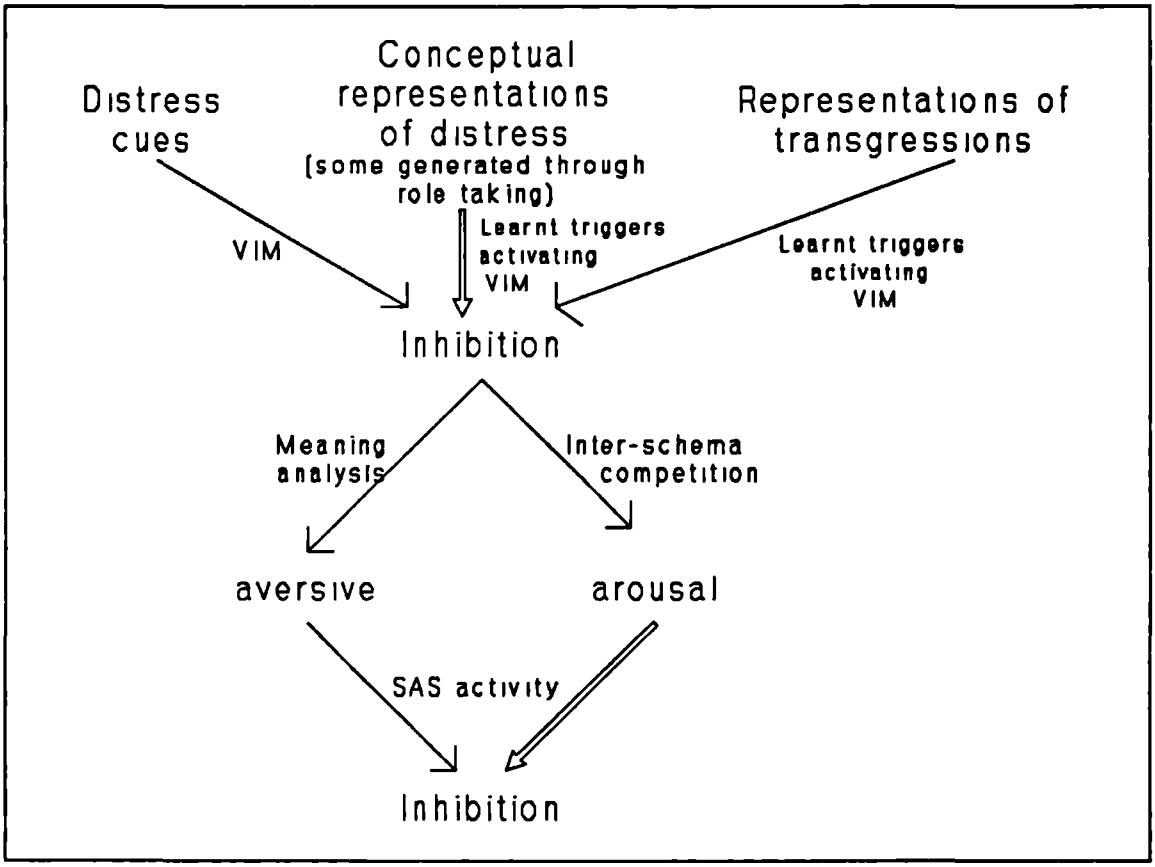
Figure 2.10. Figure 2.10(a) represents an information flow model of the operation of VIM. Figure 2.10(b) represents an information flow model of the empathic process.



In summary, the main difference between the operation of VIM as it is expressed above and the empathic process is that under the operation of VIM, the arousal is a consequence of the inhibition, while in conceptualisations of the empathic process, the arousal (i.e., empathy) causes the inhibition. However, the information flow model of VIM represented in figure 2.10(a) does not consider the influence of the learned triggers for VIM (see above and figure 2.9). Through a process of classical conditioning, VIM comes to be activated by conceptual representations of distress. I claim that some of these conceptual representations of distress will be the product of role taking. Thus, the empathic process is conceptualised within the current account as the activation of VIM by conceptual representations of distress formed though role-taking. Of course, this does not change the later stages of the information flow model. It is still the inhibition of behaviour by VIM which generates the arousal and not the other way round.

However, it is not difficult to imagine that the generation of aversive arousal (i.e, empathy) may play a role in the inhibition of violence. In section 2.5.4, I will describe how VIM may be over-ruled. The operation of VIM does not necessarily result in the inhibition of violence. In section 2.4.2, I mentioned that the role of arousal was to alert the SAS to the specifics behavioral interruption. It is expected that aversive arousal (arousal + withdrawal response through meaning analysis) should result in the inhibition of the behaviour itself. Thus, a fuller description of the relationship between the operation of VIM and the empathic process should look more like the information flow model of figure 2.11.

Figure 2.11. An information flow model of the operation of VIM incorporating the elements of the empathic process. These elements are represented in figure 2.11 by the thicker arrows. As can be seen, I claim that previous conceptualisations of the empathic process have neglected the role of inhibition in the generation of the arousal.

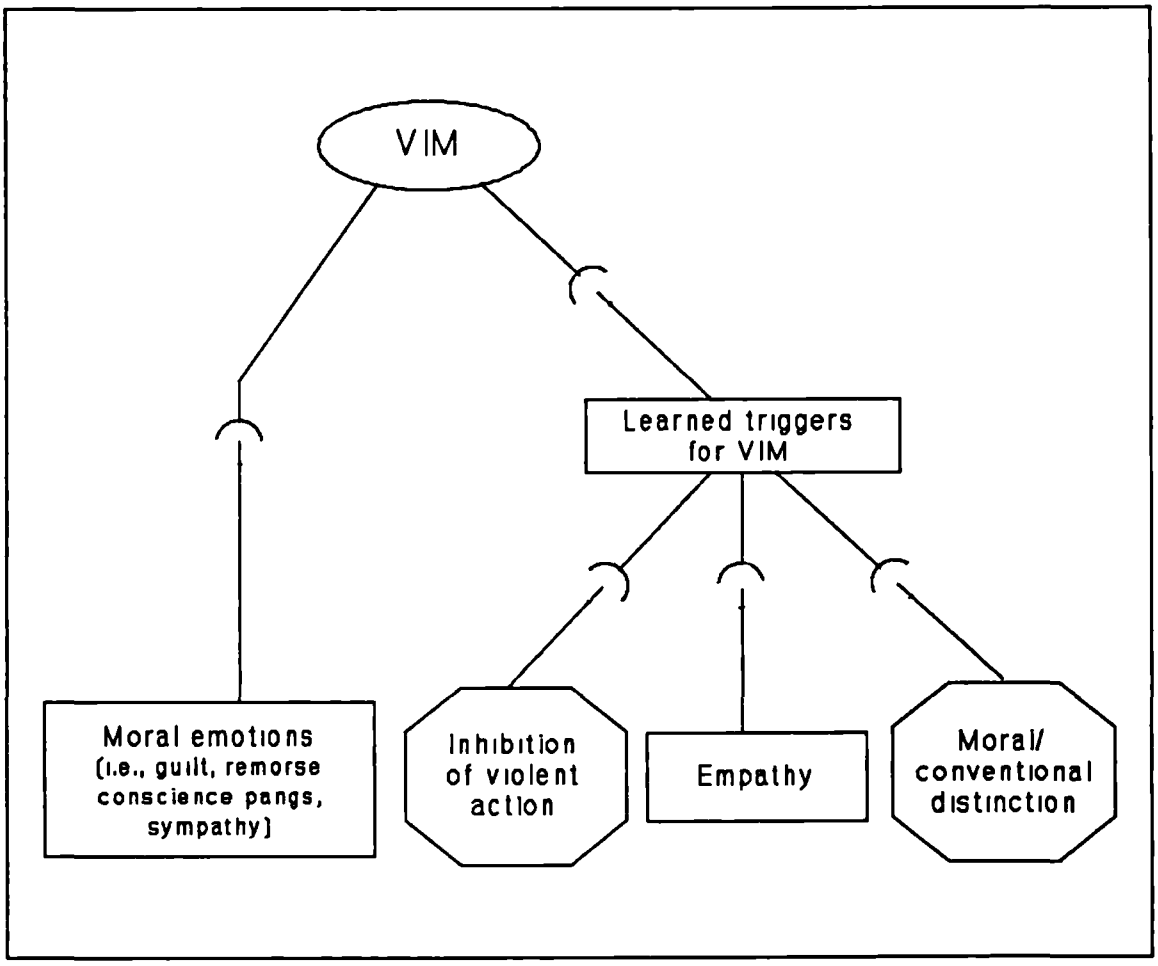


2.5.3: The developmental consequences of VIM

Figure 2.12 details a Developmental Contingency Model (DCM) for three components of morality: the inhibition of violence; the moral emotions and; the moral/ conventional distinction.

DCMs are models of developmental contingency (*see* Morton, 1986). Within a DCM, elements are identified as prerequisites for the development of other elements. For example, in figure 2.12, I show VIM as a prerequisite for the "moral emotions" and the "learned triggers for VIM". Different types of elements are symbolized distinctly in a DCM. Primary, innately specified mechanisms are symbolized by eclipses. In figure 2.12, VIM can be seen to be conceptualized as a primary mechanism. Non-primary mechanisms, and forms of representation, are symbolized by rectangles. In figure 2.12, "meaning analysis" and "empathy" are symbolized by rectangles. Behaviours are symbolized by octagons. In figure 2.12, it can be seen that the "inhibition of violent action" and the "moral/ conventional distinction" are behavioral, developmental consequences of the formation of the relevant "learned triggers for VIM".

Figure 2.12: A DCM for three components of morality: the inhibition of violence; the moral emotions and; the moral/ conventional distinction.



I have shown VIM as a prerequisite for the development of the moral emotions. The implication

of this is that individuals lacking VIM will not develop the moral emotions. The activation of VIM provides the arousal that is interpreted through meaning analysis as one of the moral emotions. Without VIM, this arousal will not be generated and the moral emotions will not develop.

VIM is also a prerequisite for its own development; i.e., the addition of "learned triggers for VIM". These additional learned triggers for VIM are learnt through repeated pairings of the US of distress cues with the CS of conceptual representations of distress/ the transgression (see figure 2.9b). Obviously, experience of the conceptual representations of distress/ the transgression are also necessary for the formation of additional triggers for VIM. This is represented in figure 2.12.

I have shown the "learned triggers for VIM" as a prerequisite for the "inhibition of violent action". The inhibition of violent action is the on-line consequence of the activation of VIM. However, this is not what I am implying in this DCM. There are at least two ways violent action can be inhibited on-line. One of these methods does not involve VIM; the SAS could deactivate the schema initiating the aggression. What I am implying here is that *developmentally* VIM, once the appropriate triggers have been learnt, will result in the inhibition of violent action. The normally developing child will be negatively reinforced by the distress cues every time he engages in any aggressive activity. Through classical conditioning this should result in even the thought of aggression being aversively reinforced; the thought of the aggression will come to trigger VIM. Hence, over time, the child will be less likely, *ceteris paribus*, to engage in violent action.[6] Thus, VIM not only inhibits aggression on-line but through the appropriate learnt triggers also developmentally reduces the probability that violence will be initiated.

The appropriate "learned triggers for VIM" are also prerequisites for empathy. As stated in section 2.5.2, empathic reactions are emotional reactions to another's distress that are defined by their representational antecedents; representations formed through role-taking. It is only after representations formed through role-taking become triggers for VIM (following classical conditioning; see above) that representations formed through role-taking can activate VIM; i.e.,

[6] Please note that I am not implicitly claiming here that children are innately specified to be aggressive. The child may engage in harmful acts through applying a motor program that has been used previously on objects. For example, the child may have pushed objects out of the way to reach other objects; the child might try to push a child out of the way to reach something.

the development of the appropriate learned triggers for VIM is a prerequisite for empathy. Hoffman (1987) has described how, over time, the range of people an individual will empathize with/ feel guilty towards broadens. At 6 years, the individual may only empathize with those he has immediately harmed. By 16, the individual may empathize with the plight of oppressed groups. I suggest that it is the development of VIM that results in these changes in who guilt is experienced towards. More specifically, I suggest that it is the expansion of the trigger data base for VIM that results in the changes in who guilt is experienced towards. At 6, VIM may only be activated by distress cues. By 16, VIM may be activated by representations of the plight of oppressed groups.

Figure 2.12 also represents my claim that the appropriate "learned triggers for VIM" are prerequisites for the moral/ conventional distinction. I suggest that representations of moral transgressions come to activate VIM. Representations of the transgression will be repeatedly paired with the distress cues that are being caused by the act. These representations of the transgression will come, through classical conditioning, to be conditioned stimuli for the activation of VIM. Conventional transgressions do not result in victims. Conventional transgressions do not result in distress cues and nor, therefore, in the activation of VIM. Representations of these transgressions cannot, therefore, become conditioned stimuli for VIM.

I claim that it is the on-line operation of VIM which determines the moral/ conventional distinction. The on-line operation of VIM results in inhibition. The body state implications of this inhibition (effectively a withdrawal) are experienced, through meaning analysis, as aversive (following Mandler's (1984) position on value). I suggest that it is this sense of aversion to the moral transgression that results in the act being judged bad. Manipulations of the transgression's context (i.e., stating that there is no rule against the transgression) will not alter the arousal of VIM in the individual by the details of the transgression. The transgression will still be judged bad. Conventional transgressions will not generate this sense of aversion. They are defined as transgressions by the presence of rules. Removal of the rule by modifying the transgression context and the transgression should no longer be judged bad.

It is interesting to recall here the findings of Davidson *et al.*, (1983) and Smetana (1985) that I reported in chapter one. Davidson *et al.*, (1983) claimed that they had shown that the child's familiarity with a transgression determines whether the transgression is judged as moral or

conventional.[7] Smetana (1985) demonstrated that the perceived consequences of a transgression could determine whether a transgression is judged as moral or conventional. Both of these findings are in line with my claim that it is the operation of VIM that is mediating the observed moral/ conventional distinction. The more familiar a moral transgression, the more likely an individual is to have paired the activation of VIM by distress cues with a representation of the transgression. Thus, following the conditioning account of the development of VIM (section 2.5.1), the more familiar a transgression the more likely it is that a representation of the transgression will have become a trigger for the activation of VIM. If VIM is activated by a representation of the transgression, that transgression will be judged as moral (see above). As regards the Smetana (1985) findings, in section 2.5.1 I claimed that pairings of conceptual representations of distress (i.e., "That person is in pain") with the activation of VIM by distress cues would result in these conceptual representations becoming triggers for the activation of VIM. I claim that if VIM is activated by a conceptual representation of distress, the transgression that resulted in the formation of that conceptual representation of distress will be judged as moral.[8]

2.5.4: Deactivating VIM

The ethologists (Eibl-Eibesfeldt, 1970; Lorenz, 1966; 1981) who claimed that humans lacked the mechanisms to control their aggression did so because of the violent nature of the species. I claimed above that humans do have functionally similar mechanisms to those possessed by many other social animals for the control of aggression. I will now explain how so much violence can be possible despite VIM.

Distress cues activate VIM. In any situation where distress cues are displayed the observer's VIM will inhibit other schemata which will, in turn, result in arousal. This means that individuals in distress cue situation should experience high levels of arousal. This does not

---

[7] However, their findings may be re-interpreted similarly to the Smetana (1985) findings; that it is saliency of the intrinsic consequences of the transgression that determines a transgression's domain; see section 1.3.2., chapter one.

[8] It should be noted here, that once conceptual representations of distress have become triggers for the activation of VIM it becomes easier to make transgressions that would not normally result in distress be judged as moral transgressions. It is only necessary to suggest that if the individual does that particular conventional transgression it will result in distress. For example "Clean your room otherwise your mother will be upset" or "Every time you swear, Jesus bleeds".

mean, however, that individual's in distress cue situations should always cease violent activity. As stated in section 2.2, behaviour can be determined either by the inter-schema competition process or through the influence of the SAS. I suggest that in those situations where an individual does not cease his aggression, the SAS is "over-ruling" VIM.

Examples of the effects of continuing aggressive activity in the face of distress cues can be found in the empirical literature. In the Milgram (1974) experiments, subjects thought that they were administering electric shocks to other individuals as punishment. These subjects showed signs of extreme distress (e.g., sweating, trembling, laughing and stuttering); (Williams, 1992). Anecdotal reports can also be found:

> "I have experienced the face of the victim. I have seen the torturer's face at close quarters. It was in a worse condition than my own bleeding, livid face. The torturer's was distorted by a kind of twitching that had nothing human about it. He was in such a state of tension that he had an expression similar to those we see on Chinese masks." (Geo Mangaskis, *Letter to Europeans*, Amnesty International Report, 1973).

I suggest that in these cases the SAS is over-ruling VIM. Orders enter the SAS as externally generated plans. These drive behaviour competing directly with VIM for the activation levels of aggression schemata. In line with this, Haritos-Fatouros (1983) found that Greek soldiers convicted on charges of torture were reported to have been obedient sons, and in the Greek army recruits were selected as potential torturers because they obeyed senseless orders without hesitation (Gibson & Haritos-Fatouros, 1986). Presumably, the SASs of these subjects were relatively more likely to "win" the competition process. Also, Milgram (1984) noted that the degree of authority pressure to continue was a significant determinant of subject "shocking" (subjects were significantly less likely to cease "shocks" if the experimenter instructed them to continue). Presumably, providing the SAS of subjects with orders makes it more likely that it will initiate the competition process.

I also suggest that in these situations of SAS over-ruling the arousal is re-attributed; i.e., it is not attributed as being caused by the distress cues the victim is exhibiting. Staub (1990) notes that Greek torturers were selected from recruits from families who were known to oppose Communism. It seems possible that the arousal might be re-attributed as hate for the torture victim; through meaning analysis the individual labels the aversive arousal as "hate for the victim" and not "hate for the activity being done to the victim". Perhaps, re-attribution may also result in the aggressor enjoying the aggression. Following Mandler claims (e.g., 1984; see

section 2.31) that value judgements are determined by a process of "meaning analysis" on the approach/ avoidance tendencies the subject is activating at the time of the event. If the individual initiates approach tendencies during the violence, he will enjoy the violence. Possibly, "low power" individuals (those worried about their lack of status) may enjoy attacking "high power" targets because of approach tendencies initiated by the possibility of the close proximity to the "high power" target.

As stated above, distress cues activate VIM. A simple way of de-activating VIM is to make sure that the attacker does not experience the distress cues of the victim. "If there is any threat that the victim might, for an instant, touch the torturers humanity, then a black bag is thrown over his head or he is strapped to other end of the machine" (Vietnam Veteran, Amnesty Report, 1973). In line with this, Milgram (1974) found that the proximity of the subject to the victim was a significant determinant of tendency to shock (subjects were significantly more likely to cease "shocks" if the "victim" was in the same room, than if they would only be heard, than if they could neither be seen or heard).

Once conceptual representations of distress and representations of transgression have, through classical conditioning, become triggers for VIM, VIM does not need the presence of distress cues to become activated. Once conceptual representations of distress (i.e., those formed through role-taking) have become triggers for VIM, VIM may be activated by representations of the transgressions and conceptual representations of distress. It is possible to prevent the activation of VIM, in situations where the distress cues may not be present, by preventing the activation of these learnt trigger nodes. Transgressions may come to be re-represented by euphemisms which will not activate the learnt trigger nodes. For example, Greek torturers referred to a beating with fists as a "tea party" and a beating with fists and clubs as a "tea party with toast" (Haritos-Fatouros, 1988). Lifton (1986) notes that Nazi doctors reconstructed morality to the point where they were able to discuss ways of carrying out genocide "humanely", when genocide was referred to as the "Final Solution". Human beings are dehumanized. The Amnesty Report (1973) notes that a precondition for being a torturer is "having a world view, however crude, that divides men into the torturable and the non-torturable". Dehumanized people will no longer activate the learnt trigger nodes:

> "When you shot at someone you didn't think you were shooting at a human. They wee a Gook or a Commie and it was okay." (Vietnam Veteran, Amnesty Report, 1973).

Similarly, the Argentineans became "Argies" for the duration of the Falklands War and the supporters of the two Sheffield football teams refer to supporters of rival team as "pigs" (Armstrong, 1992). In addition to preventing the activation of the learnt nodes, I suggest that dehumanization reduces the likelihood of forming the representations through role-taking that might excite the learnt nodes. A Vietnam Veteran writes on the experience of killing a child:

> My first reaction was, I guess, you would call normal. It would be horror, pain, and when I realised that I caught myself immediately and said 'No, you can't do that', because you develop a shell when you are in the military. They brainwash you. They take all the humanness out of you and you develop this crust which enables you to survive ... And if you let that protective shell down, even for a second- it's the difference between you flipping out or managing to make it through." (Amnesty Report, 1973).

I also suggest that dehumanizing the enemy prevents the acts that are committed against them being represented as transgressions. It is possible to kill vermin. Killing vermin is not represented as a transgression. Representing the enemy as vermin prevents their killing to be represented as a transgression. If the act is not represented as a transgression the learnt nodes and consequently VIM will not be activated.

Dehumanization also has implications both for those actually involved in the violence and those who hear about it. While dehumanization will not prevent the activation of VIM at the time the act is committed (if there are distress cues, VIM will fire), dehumanization will reduce the likelihood that the individual will review the acts thus re-activating VIM through the learnt nodes. Dehumanisation will also prevent the activation of VIM through learnt nodes by those who only receive reports of the aggression (e.g., those "at home" during a war or those hearing what their criminal or politically motivated colleagues). Without the learnt nodes being activated these people will not be aversively aroused by the victims's plight.

2.6: Conclusions

In this chapter, I described the Supervisory Attentional System Framework; a framework for the mental control of action. I described two models of emotion [Mandler's (e.g., 1984) and Oatley & Johnson-Laird's (1987)] and developed a third model of emotion within the SASF. I then went on to describe the characteristics and development of a proposed Violence Inhibition Mechanism (VIM) and to explore, through Developmental Contingency Modelling, its

developmental consequences. The DCM makes the three predictions that have been tested in this thesis:

(1) Individuals who possess VIM will make the moral/ conventional distinction.

(2) Individuals who lack VIM will fail to make the moral/ conventional distinction.

(3) Individuals who lack VIM will lack the moral emotions.

Prediction one is tested in chapter three, prediction two in chapter four and prediction three in chapter seven.

Chapter Three

Autism and Morality

3.1: Introduction

Chapter two developed a theory of the development of morality which made several testable predictions. One of these predictions was that any individual who is capable of the proposed aversive reaction to distress cues will develop morality; this individual will differentiate between moral and conventional transgressions. The theory therefore predicts that autistic children, assuming they are capable of generating this aversive reaction to distress cues, will distinguish in their judgements moral and conventional transgressions; they should have developed morality in at least this form. However, previous theories would not make this prediction. For example, Turiel (1977) claims that role-taking is one of the techniques the child has for the construction of concepts.[1] Since autistics cannot role-take (see below) they should not, according to Turiel, make the moral/ conventional distinction.

In this chapter, I will initially describe the history of the clinical diagnosis of autism (section 3.1.1). Following this, I will briefly review the literature describing the cognitive deficit underlying autism (section 3.1.2). Finally, in section 3.1.3, I will consider the relationship of this cognitive deficit (the lack of a Theory of Mind Mechanism) to the development of morality.

3.1.1: An introduction to autism:

Autism is a developmental disorder affecting between 4 and 10 in every 10,000 children born (Lotter, 1966; Bryson, Clark & Smith, 1988; Steffenberg & Gillberg, 1986). The disorder was first named and described by Kanner (1943). His description focused on the behaviour of the individual. His essential and defining symptoms were the child's "autistic aloneness" and "obsessive desire for the preservation of sameness" (Kanner & Eisenberg, 1956).

Kanner's (1943) description was based on a limited number of cases that were referred to his clinic. No examination was made to see if some of the features he observed were an indication of problems additional to the child's autism. Thus, while Wing and Wing (1971) found that

[1] Indeed, role-taking may be the only one of Turiel's techniques that could give rise to the moral/ conventional distinction (see below and chapter one).

more than 80% of autistic children in their sample showed preference for the proximal senses (a feature noted by Kanner), this preference was also seen in 87% of partially blind and deaf children, 47% of subjects with Down Syndrome and 28% of normal children.

In addition, the behavioral manifestations of autism can be seen to vary across, and even within, individuals, according to age and intellectual ability. Some children with autism avoid social contact, like Kanner's cases, others are merely passive, or even actively social in a peculiar fashion (Wing & Gold, 1979; Wing, 1988). However, it is apparent that there are three core impairments in autism: problems with socialisation; problems with communication and; problems with imagination. These three core impairments have become known as Wing's triad (Wing, 1981) and they were identified through epidemiological studies (e.g., Wing & Gould, 1979).

Wing and Gould (1979) assessed all the children living in the Camberwell area at the time of the study (N = 35,000). All children known to the social, educational or health services were screened (N = 914). Children were then selected from this group if they showed one, or more, of the following: social impairment, verbal or nonverbal language impairment, repetitive/ stereotyped activities. The resulting 132 children (aged 2 to 18) were observed and given medical and psychological tests, and their carers were interviewed with the Handicap, Behaviour and Skills Schedule (Wing & Gould, 1978). The group was then divided on the basis of social behaviour into 58 children with appropriate social interaction (for their MA) and 74 socially impaired subjects (of whom 17 were classically autistic, by Kanner and Eisenberg's (1956) criteria of social aloofness and elaborate routines). The groups did not differ significantly in age, but there were more males in the socially impaired group than in the sociable group.

Significant differences were observed in the communicative and play behaviours of the two groups; 90% of the sociably impaired (versus 50% of the sociable group) were either mute or echolalic at the time of parental interview, and 97% of the socially impaired group (versus 24% of the sociable group) showed no or only repetitive symbolic play. The 24% of the sociable group who did not show symbolic play all had a language comprehension of less than 20 months - a mental age below which pretence would not be expected, since normal children only manifest this ability in the second year of life. By contrast, the socially impaired above 20 months of age still showed a poverty in symbolic play. Wing and Gould concluded that:

> "all the children with social impairments had repetitive stereotyped behaviour and almost all had absence or abnormalities of language and symbolic activities. Thus the study

> showed a marked tendency for these problems to occur together" (Wing & Gould, 1979; p. 25).

This association between deficits in socialisation, communication and imagination was also found in a group of 761 adults in a mental handicap hospital; abnormal speech was shown by 75 % of those with social impairment, versus 14 percent of those showing social interaction appropriate for their mental age. Lack of symbolic activity was found in 73% of the socially impaired group, and only 8% of the sociable group. The association also emerges if the Camberwell sample is divided on the basis of types of play shown (Wing, Gould, Yeates & Brierley, 1977) rather than social understanding.

Today, the diagnosis of autism in both DSM-III-R (American Psychological Association, 1987) and ICD-10 (World Health Organisation, 1987) is based on three fundamental impairments which capture Wing's triad:

(1) Qualitative impairment in reciprocal social interaction.

(2) Qualitative impairment in verbal and nonverbal communication and in imaginative activity.

(3) Marked restricted repertoire of activities and interests.

3.1.2: The cognitive deficit underlying autism

It has been proposed that autistics lack a Theory of Mind Mechanism (ToMM); *see*, Leslie, 1987; 1988. This mechanism allows the representation of the mental states of others; *meta-representations*. It has been suggested that the autistic child cannot form meta-representations (e.g., Frith, 1989; Leslie, 1987; 1988) and that it is this inability that gives rise to Wing's triad of symptoms (Frith, Morton, & Leslie, 1991). The social impairment of autistics would follow directly from their not being able to understand the actions of others as resulting from distinct beliefs and desires. The characteristic communicative impairments of autistics would follow from their inability to represent the intentions of others and to see beyond the literal interpretation of what others say. The third impairment, problems with imagination, reveals itself most dramatically through pretend play. A child without ToMM should be unable to engage in pretend play. Initially, pretend play is always shared. Thus, it is only possible if the

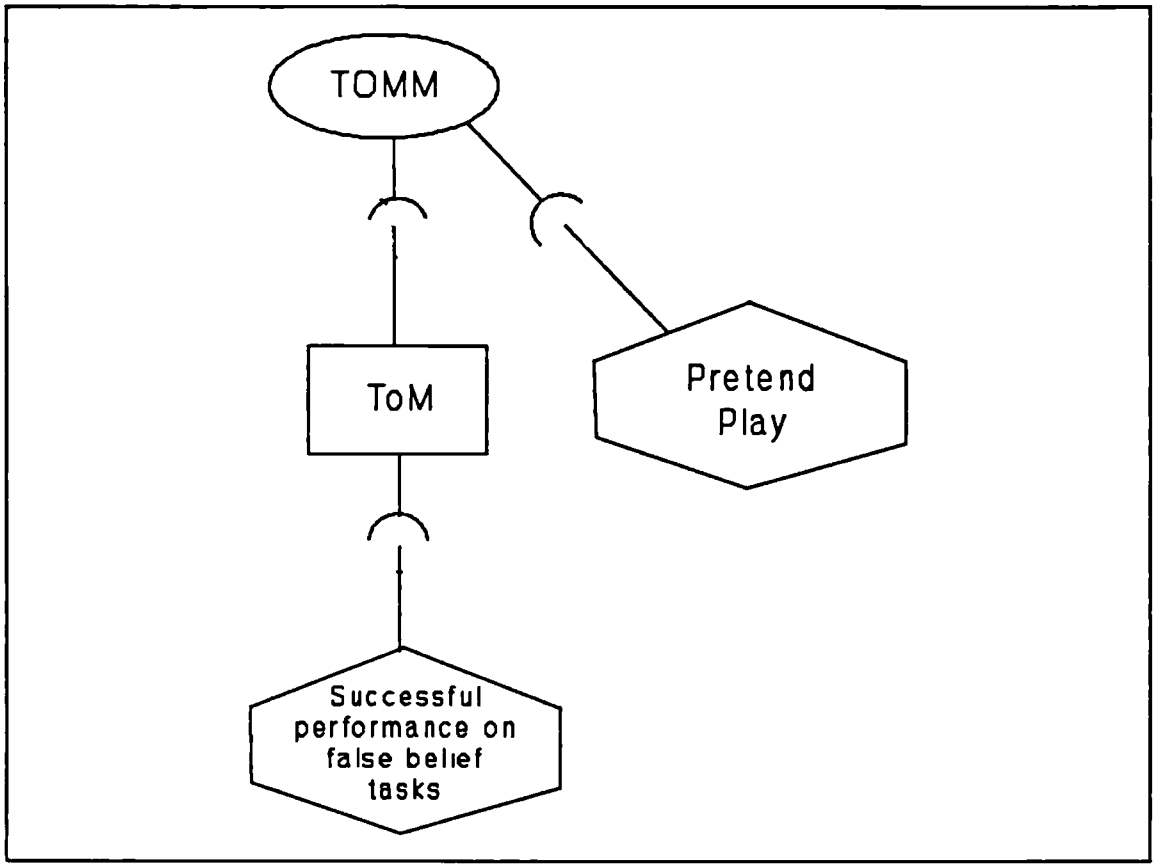
child is able to understand the mental states of the person sharing the pretend (*see* Leslie, 1987).[2]

ToMM is a prerequisite [see figure 3.1 for a Developmental Contingency Model (DCM) of ToMM] for the development of a "Theory of Mind". While ToMM allows the representation of the mental states of others, the "Theory of Mind" provides knowledge of what these mental states might be. Within the model developed in the Supervisory Attentional System framework (SASF) that was described in chapter two, ToMM is assumed to be a mechanism in the SAS which "decouples" primary representations to form meta-representations (*see* Leslie, 1987). The "Theory of Mind" is assumed to be a schema within contingency scheduling which, when activated, passes representations of the thoughts of others to the relevant awareness buffers in the Supervisory Attentional System (SAS). A "Theory of Mind" is assumed to calculate representations of the mental states of others given particular environmental circumstances transmitted to the trigger data base. Without ToMM the products of a "Theory of Mind" cannot be expressed. It is assumed that the development of a "Theory of Mind" is a prerequisite for successful performance on first order false belief tasks; e.g., the Sally-Anne task. First order false belief tasks involve the child making attributions of the form "X thinks Y" when Y is not the case; i.e., "Sally thinks the sweet is in the basket" when the sweet is in the box. The Sally-Anne task involves a doll who puts an object into one location and who is then absent when the object is unexpectedly transferred to another location. Understanding the doll's resulting first order false belief about the object's location is indicated if the subject predicts that on her return the doll will look for the object in the original, now empty location (Wimmer & Perner, 1983). Baron-Cohen, Leslie and Frith (1985) tested 20 autistic children with MA's of over four years (mean nonverbal MA 9;3, VMA 5:5) on the Sally-Ann task. 80% of these children failed to appreciate Sally's false belief. They did not predict that Sally would look in the original location. In contrast, 86% of Down Syndrome children of lower mental age (nonverbal 5:11, VMA = 2:11) succeeded on the task. They predicted that Sally would look in the original location. This finding has been consistently replicated, using real people instead of toys, using a "think" question ("Where does Sally think the object is?") rather than a "look" question ("Where will Sally look for the object?"), and using a control group of specifically-language impaired children to rule out a language deficit explanation (Leslie and Frith, 1988; Nunez & Riviere, 1990; Perner *et al.*, 1989; Reed & Peterson, 1990; Riviere & Castallanos, 1988).

---

[2] Wulff (1985), reviewing the literature, concluded that autistic children do not show pretend play.

Figure 3.1: A DCM of ToMM.



Autistic children have been shown to fail other first order false belief tasks; *see*, for a review of this literature, Happé (1991). For example, Dawson and Fernald (1987) have shown autistics to be impaired in "conceptual perspective taking"; they cannot choose gifts that are appropriate for different people. Perner *et al.*, (1989) have shown that autistics fail the "smarties" task; another "false belief" task. In this task the child is shown a smarties tube and is asked what is inside. It is then demonstrated to the child that there is a pencil inside and not sweets. The child was then asked what another child from their class (not previously tested) would say was inside the packet. Understanding the other child's resulting false belief about the packet's contents was indicated if subjects predicted that the other child would say "sweets".

A "Theory of Mind" is also a prerequisite for successful performance on second order false belief tasks. Second order false belief tasks involve the child making attributions of the form "A thinks B thinks Y" when Y is incorrect; i.e., "Mary thinks John thinks the ice-cream van is in the park" when the ice-cream van is near the church. The classic second order false belief task is that of Perner & Wimmer (1985). John is told in the presence of Mary that the ice-cream

van will be in the park all day. John goes home to get the money for an ice-cream. The ice-cream man tells Mary that he is going to the church to sell his ice-creams. On the way, the ice-cream man sees John and tells him where he is going. Mary goes to John's house and asks his mum where he is. John's mum tells her that he is buying an ice cream. The subject is asked "Where does Mary think that John has gone to buy his ice-cream?". Understanding Mary's resulting second-order false belief about John's belief was indicated if subjects answered that Mary thought that John was in the park. Baron-Cohen (1989) observed that even those autistics who pass first-order false belief tasks fail this second-order false belief task.

There have been some claims that autistic children do not suffer from a difficulty in the formation of representations of the mental states of others. Researchers have suggested that the failure of autistics on false belief tasks is due to pragmatic difficulties with the "look" question (Eisenmajer & Prior, 1991), to perceptual salience of the real location (Russell, Mauthner, Sharpe & Tidswell, 1991; Hughes & Russell, 1990), or to a lack of motivation to succeed (De Gelder, 1987). Most such criticisms do not address the fact that autistic children fail a whole array of false belief tasks with very different controls and methodologies while they succeed on a variety of tasks that do not require the processing of representations of false belief. For example, Sodian and Frith (1992) examined the relative ability of autistic children to keep a sweet from a puppet competitor by deception and by sabotage. By contrasting these two conditions it was possible to rule out lack of motivation or failure to comprehend the instructions as explanations of failure on the deception task. The only difference between the two conditions was that deception (lying or pointing to the empty location) manipulated the competitor's beliefs, while sabotage (locking the box in which the sweet was hidden) manipulated his behaviour. While autistic children proved competent at sabotage they were largely incapable of deception.

Another challenge to the "Theory of Mind" theory of autism is the observation that a minority of autistic children pass "Theory of Mind" tasks. For example, 20% of the autistic children in the Baron-Cohen *et al.*, (1985) study passed the Sally-Anne task. However, as stated above Baron-Cohen (1989) found that even those autistics who passed the Sally-Anne task failed the second-order "ice cream task".

However, Bowler (1989) found that a group of "Asperger's Syndrome" subjects (who can be defined as a very high functioning group of autistics) do just as well on second order false belief tasks as normal controls. In contrast, Ozonoff, Rogers & Pennington (1991) have shown that

the high functioning autistic but not "Asperger's Syndrome" subjects fail second order false belief tasks.

Happé (1991) has extensively analyzed the functioning of these high ability autistics. She has demonstrated that three groups of autistics can be reliably distinguished. She divided autistic subjects up into three groups according to their ability on false belief tasks: the first group passed none of the false belief tasks; the second group passed first order false belief tasks (e.g., the Sally-Anne task and the Smarties task) and; the third group who passed first and second order false belief tasks (the ice cream van task). Happé gave these subjects the "strange stories" task. This involved the experimenter reading the subject short stories detailing social situations which needed the calculation of mental states to be understood. These stories were seen as more naturalistic tests of "Theory of Mind" ability than the classic false belief tasks such as Sally-Anne. Happé found that level of ability on the classic false belief tasks predicted score on the "strange stories" task. There was a significant difference between the scores of the three groups on this task (no ToM autistics scoring very low, second order autistics scoring relatively very high).

Happé (1991) also looked at the communicative ability of these three groups of autistic subjects, generating predictions from the linguistic analysis of Sperber and Wilson's (1986) relevance theory. Sperber and Wilson analyzed the cognitive requirements necessary for the understanding of various linguistic structures; similes, metaphors and irony. Working within their framework, Happé proposed that similes can be understood at a purely literal level. For the individual to understand "Peter is like a wolf" it is only necessary for that individual to identify the points of similarity between their concept of wolf and their concept of Peter. They proposed that the understanding of metaphors ("Peter is a wolf"), however, necessitated some understanding of intentions, since metaphors cannot be understood literally. Metaphors can be seen as analogous to false belief statements; to understand them it is necessary to calculate the mental state of the speaker. Finally, Happé proposed that the understanding of irony required second order meta-representation (a thought about a thought). Irony ("What a lovely day for a picnic" when it is raining) involves mentioning a possible thought and expressing an attitude to that thought. Happé (1991) predicted from this analysis: first, that autistics with no ability on "Theory of Mind" tasks would be able to understand similes but be significantly impaired in their understanding of metaphor and irony; second, that autistics who could pass false belief tasks but not second order tasks would understand similes and metaphors but not irony and; third, that

autistics who could even pass second order tasks would understand similes, metaphors and irony. These predictions were confirmed.

It seems, therefore, that subgroups of autistics can be distinguished according to their ability on "Theory of Mind" tasks and that their ability on these tasks is directly correlated with their communicative competence. It is, however, unclear, whether these groups differ in the possession of a "Theory of Mind" or in the possession of alternative cognitive functions that allow them to pass false belief tasks and understand non-literal utterances. It is unclear whether the DCM of figure 3.1 represents the development of these children or whether their development should be represented by a different model.

Baron-Cohen (1988) suggested that autistic people are merely slower in their acquisition of a "Theory of Mind", and that it is therefore no surprise that a few autistic people should manage to pass these tests eventually. According to Baron-Cohen, there is a developmental delay in the acquisition of a "Theory of Mind" in autistics. In accordance with, rather than proof of, this, Frith *et al.*, (1991) found that of 42 autistic subjects tested on false belief tasks, those who passed were (with only one exception) over 11.5 years old and of a VMA of over 5.5 years. It must be remembered, however, that there are equally advanced autistic children who fail these tasks.

Frith *et al.*, (1991), however, interpret their findings not as evidence for delay, but as evidence that autistics can pass "Theory of Mind" tasks through the use of non-"Theory of Mind" strategies (Frith *et al.*, 1991). They suggest the "passer's" superior IQ and age are signs that the "passers" have managed to hack out a solution to the puzzle only thanks to experience, and using generalized problem solving skills (Frith *et al.*, 1991). It also might be expected, if the passing autistics are using non-"Theory of Mind" strategies, that their ability to pass "Theory of Mind" tasks might be related to some measure of general problem solving skill while this would not be the case in control groups. Recent work by Riviere and colleagues supports this idea. Riviere and Costellanos (1988) found that good "Theory of Mind" performance was highly correlated with success on a test of operational thinking in autistic subjects. No such correlation between these tasks was found in normal 3- to 5- year old children (Nunez & Riviere, 1990).

3.1.3: Autism and morality

Chapter two explored a theory of the development of morality. This account suggested that the generation of morality (as evidenced by the child making a distinction between moral and conventional transgressions) was dependent on the child possessing a Violence Inhibition Mechanism (VIM). It was postulated that VIM was innately specified; that it was the functional equivalent in humans of the system in animals that inhibits aggression following species specific distress/ submission cues.

In chapter two I developed a DCM of VIM (see figure 2.12). That model represented my claim that VIM is the only basic mechanism prerequisite for the moral/ conventional distinction. In other words, my theory would predict that any individual who has formed VIM will distinguish in their judgements between moral and conventional transgressions. There is no reason to believe that autistic subjects lack VIM. There is no reason to believe that the autistic would not be able to form representations of the event (e.g., "presence of victim"). These would be *primary representations* which, Leslie (1987) has claimed, the autistic can form. Referring back to the Supervisory Attentional System Framework (SASF), VIM is a schema within contingency scheduling. ToMM, in all probability, is a Supervisory Attentional System (SAS) mechanism that operates on representations in the awareness buffers. Thus, according to the DCM presented in chapter two (figure 2.12), the autistic should distinguish in their judgements between moral and conventional transgressions. This is despite their lack of ToMM.

However, other accounts would suggest that ToMM is a prerequisite for the moral/ conventional distinction. Several theorists (e.g., Colby, Kohlberg, & Kaufman, 1989; Piaget, 1932; Turiel, 1977) have explicitly linked the ability to role-take, with the generation of morality. Role-taking is defined as the "imaginative transposing of oneself into the thinking and acting of another" (Feshbach, 1978). From this definition, it would seem that role-taking involves the formation of mental states and thus the activity of ToMM. It is only possible to role-take if it is possible for the system to represent the thoughts of others. Indeed, the development of a "Theory of Mind" (see figure 3.2) must also be a prerequisite for the ability to role-take. As stated above, a "Theory of Mind" calculates representations of the mental states of others given particular environmental circumstances. Role-taking, effectively the specifying of the environmental circumstances to drive the calculation process, obviously requires the calculation ability.

Role-taking and the operation of VIM are very different activities; they differ in their activation conditions, their operation and in their outputs. In particular, role-taking and the operation of VIM differ in their degree of automaticity. VIM is activated automatically by representations of distress cues, processed by the sensory systems, stimulating the trigger in the trigger data base responsible for the activation of VIM. Role-taking, on the other hand, by its very nature (the imagining of another's thoughts and feelings), cannot be an automatic process. It must involve planning; the stimuli relevant to the goal of the role-taking must be identified and appropriate schemata for representation activated. In terms of the SASF, the SAS must be involved.

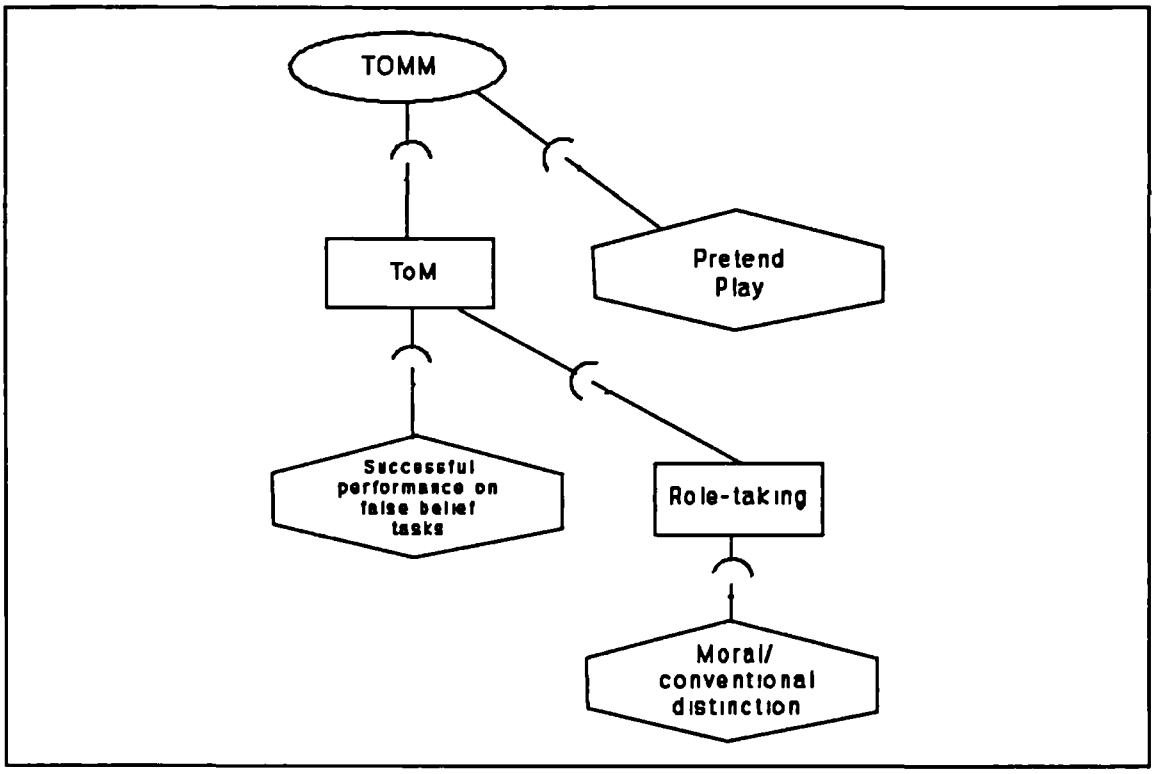
Role-taking and VIM differ in their operation. VIM, as it is conceptualised in chapter two, should be thought of as one schema which inhibits behaviour in the context of distress cues. Role-taking, on the other hand, is a collective term describing any processing operation that results in a representation of another's thoughts or feelings. For example, role-taking might involve either imagining the thoughts and feelings of oneself in that situation or an actual attempt to imagine what another is thinking and feeling.

Role-taking and VIM also differ in their outputs. VIM principally results in a behavioral output; the inhibition of behaviour. As reported in chapter two, behavioral interruptions generate arousal (*see*, Mandler, 1984; 1987; 1991). Thus, a second output of VIM activation is arousal. However, other than the generation of the arousal, the operation of VIM has no representational consequences. Role-taking, on the other hand, results in representations of the thoughts or feelings of the other person.

Piaget argued that role-taking was one of the factors that was necessary for the formation of autonomous morality. Kohlberg suggested that an ability to role-take was necessary if the child was to reach a post-conventional level of morality. Turiel claimed that role-taking was one of the methods available to the child for the construction of moral and conventional concepts. Claims that role-taking is involved in the development of the moral/ conventional distinction allow several other predictions. First, as is represented in the DCM of figure 3.2, role-taking might be a prerequisite for the moral/ conventional distinction. If this is the case, it would be predicted that zero-order autistics (those autistics who fail all false belief tasks and who can be concluded to be lacking a "Theory of Mind") will not make a distinction between moral and conventional transgressions. Predicting the behaviour of first- and second-order autistics is more complicated. If the possession of a "Theory of Mind" is a prerequisite for the formation of a

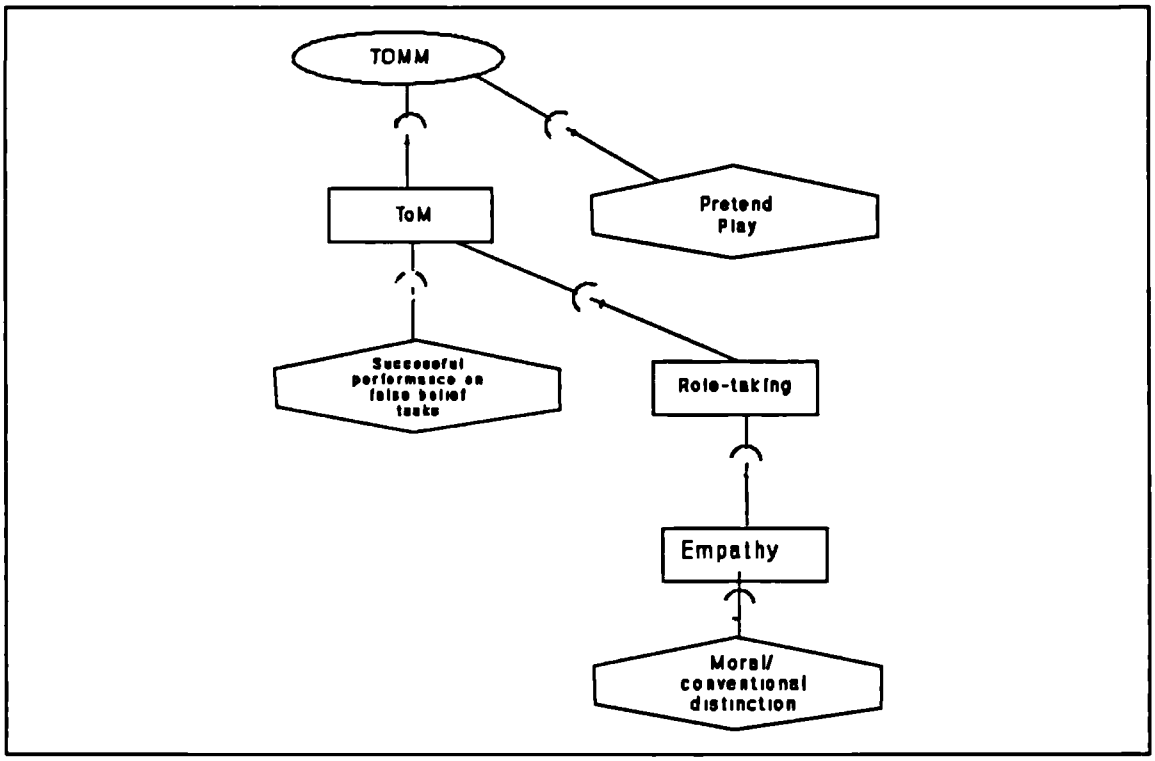
moral/ conventional distinction, and the developmental lag position of Baron-Cohen (1988) is correct, then it would be predicted that first- and second- order autistics would make the moral/ conventional distinction. Figure 3.2 represents a DCM of this situation.

Figure 3.2: A possible DCM of the role of ToMM in the development of the moral/ conventional distinction.



Examining the moral/ conventional distinction of autistic subjects also allows conclusions to be drawn about the importance of empathy for the development of the distinction. Empathy is defined as "an affective response more appropriate to someone else's situation than to one's own" (Hoffman, 1987; p. 48); see section 2.5.2. Empathy has also been implicated in the development of morality. Hoffman (1987) has suggested that empathy may be involved in the selection of moral principles. Kagan has suggested that empathic reactions to moral, but not conventional, transgressions might be one of the causes of the moral/ conventional distinction. In figure 3.3, I present a DCM showing empathy as a prerequisite for the moral/ conventional distinction.

Figure 3.3: A DCM showing empathy as a prerequisite for the moral/ conventional distinction.



The performance of autistic children on the moral/ conventional distinction may help to determine the role of empathy because, according to the dominant conceptualisations of the empathic process, it is representations formed through role taking that activate the empathic process. As Batson states: "Perspective taking is the psychological variable most often assumed to be the antecedent of specifically empathic reactions to another's distress" (Batson, Fultz, & Schoenrade, 1987; p. 172). Autistic children cannot role take and therefore, from this conceptualisation, cannot experience empathy. Thus if, as was predicted by the DCM represented in chapter two (figure 2.12), even those autistic children who fail first order false belief tasks make a moral/ conventional distinction, empathy cannot be a prerequisite for the development of the distinction. These children cannot role take, therefore they cannot experience empathy, therefore if they make the distinction empathy cannot be a prerequisite for its development. However, if these children do not make the moral/ conventional distinction it may indicate that either role-taking or empathy are prerequisites for the moral/ conventional distinction. This particular study could not distinguish which would be the prerequisite in these circumstances.

Because of the possibility that some autistics do actually possess a "Theory of Mind", as suggested by Baron-Cohen (1988) and Happé (1991), it was decided to divide the autistics into three groups. Group one would consist of autistics who failed all "Theory of Mind" tasks. Group two would consist of autistics who passed first order tasks (Sally-Anne and smarties) but not second order tasks (the ice cream van task). Group three would consist of autistics who passed all "Theory of Mind" tasks. Given the claims of Baron-Cohen and Happé, it was felt that it was only safe to consider group one as autistics who definitely lacked a "Theory of Mind". Groups two and three were included to explore the relationship between growing sophistication on the "Theory of Mind" tasks and the nature of the moral/ conventional distinction.

3.2: Method

3.2.1: Design

The experiment involved a 5 x 2 repeated measures factorial design. The independent variables were the 5 different subject groups (one group of normal children, a group of Minimum Learning Difficulty (MLD) children and three groups of autistics selected according to their level of performance on "Theory of Mind" tasks) and the two different types of story (moral and conventional). The dependent variable was the subjects responses to the questions about the transgression situations.

3.2.2: Subjects

25 children and adults with autism took part in this study. These subjects were contacted either via special schools for autistic children or via a clinician who knew of autistics in the community. All of the subjects had received a diagnosis of autism according to DSM-III-R criteria. These subjects were divided into three groups on the basis of their performance on a battery of "Theory of Mind" tasks. The first group of subjects (N = 10) did not pass any of the "Theory of Mind" tasks. The second group of subjects (N = 10) passed both false belief tasks but failed the second order task. The third group of subjects (N = 5) passed all three tasks. There were two control groups: one group of normally developing children (N = 10) and; one group of MLD children (N = 10) and two groups of normal children differing in age. All

subjects (other than the two normal groups and one 1st order and 2nd order autistic[3]) were given the British Picture Vocabulary Scale (BPVS). Full subject characteristics are shown in table 4.1

The normally developing children and the MLD children were controls for the *no ToM* autistics. IQ was not assessed for the normally developing children and MA was assumed to approximate CA. Because autism is not a common disorder and few autistic children pass false belief tasks it was not possible to control for age and IQ across autistic groups within the time confines of this study.

Table 3.1: Characteristics of subjects.

Group	N	Age	VMA	VIQ
Normally developing children	10	8.31	--	--
MLD children	10	11.08	8.01	68.20
No ToM autistics	10	11.68	8.41	72.00
1st order autistics	10	14.63	--	79.00
2nd order autistics	5	20.03	--	93.00

3.2.3: Materials

Three tasks were used to assess the "Theory of Mind" ability of the autistic subjects; the Sally-Anne task of Baron-Cohen *et al.*, (1985), the smarties task of Perner *et al.*, 1989 and the ice-cream task of Perner and Wimmer (1985). For the Sally-Anne task a set of three small cardboard boxes of different sizes, shapes and colours served as hiding places, a one pence piece was the object hidden and Sally-Anne were Playmobil characters. For the smarties task, a smartie tube containing a pen top was used. For the ice-cream task, the village was a cardboard scene, with locations (church, park and John's house) drawn in and labelled.

The stories used to measure the moral/ conventional distinction were all taken from the literature. The four moral stories involved a child hitting another child, a child pulling the hair of another

[3] These two autistic subjects were given the WAIS.

child and the victim cries, a child smashing a piano and a child breaking the swing in the playground. The four conventional stories involved a boy child wearing a skirt, two children talking in class, a child walking out of the classroom without permission and a child who stops paying attention to the lesson and turns their back on the teacher. All props to aid story telling were Playmobil. During the stories, the adult teacher was represented by an "adult" playmobil character and the children were represented by "child" playmobil characters. Other props included a piano, a spade and a see-saw.

Subjects responses to questions were recorded on standard scoring sheets.

3.2.4: Procedure

Subjects were tested in a quiet room. Story props were always spread out on a table in front of the subjects. Autistic subjects were given the "Theory of Mind" battery and the BPVS before their moral/ conventional distinction was ascertained. 15 of the subjects received these tasks from the present author. The other subjects were tested by different experimenters. The procedures described by Baron-Cohen *et al.*, (1985) for the Sally-Anne task, by Perner *et al.*, (1989) for the smarties task and by Perner and Wimmer (1985) for the ice-cream task were faithfully followed.

The procedure for assessing the subject's moral/ conventional distinction followed the classic paradigm. The subject was told a "story" using the playmobil characters. This study would involve a rule (either moral or conventional) being broken. The order of presentation of the various transgressions was randomized across subjects. At the end of the story, the characters would be left in their final resting places (e.g., "talking" to each other or hitting each other). After the story had been told the subject was asked two questions:

(1) Was it O.K. for [*the story character*] to do [*the transgression*]? (ascertaining the children's judgement of the *permissibility* of the act)

(2) Was it bad for [*the story character*] to do [*the transgression*]? (ascertaining the children's judgement of the *seriousness* of the act)

The playmobil characters were then placed back in their original positions and the subject was

told:

> "Now what if the teacher said before the lesson, before X did Y, that "At this school anybody can Y if they want to. Anybody can Y."

The playmobil teacher character was moved as he spoke and a different tone of voice was attempted by the experimenter. The child was then asked a final question:

(3) Would it be O.K. for X to do Y if the teacher says X can? (ascertaining the children's judgement of *authority jurisdiction*[4])

Only yes/ no responses were required. All responses were recorded by hand on a standard scoring sheet.

3.2.5: Scoring procedure

The scoring procedure followed that commonly used in the literature (e.g., Smetana, 1981; Smetana & Braeges, 1990). The answers to all three questions were scored categorically. *Yes* responses were assigned a score of 0, and *no* (*not OK*) responses a score of 1.

3.3: Results

Three 2(Domain) x 5(Group) ANOVAs were performed on responses for each of the three criterion questions. Because all responses were expressed as proportions, arcsine transformations were performed to correct for non-normality (Winer, 1971). There were main effects for domain for *permissibility*, $F(1, 89) = 10.54$, $p < 0.005$; *seriousness*, $F(1, 89) = 4.69$, $p < 0.05$; and *authority jurisdiction*, $F(1, 89) = 62.25$, $p < 0.001$). There were no significant effects for group for or any significant interactions.

[4] The *authority jurisdiction* question was chosen in preference to any of the other modifiability questions (see chapter one) because this question sets up a concrete setting (the teacher saying this). It does not ask the child to imagine a country where there was no rule or just what if there was no rule. It was felt that these questions might be too abstract for the autistic subjects.

Table 3.2: The means and standard deviations of moral and conventional judgements for each of the criterion judgements for each of the subject groups.

	Criterion Judgements					
	Permissibility		Seriousness		Authority Jurisdiction	
	M	C	M	C	M	C
Normal developing	0.97 (0.18)	0.91 (0.29)	0.03 (0.18)	0.28 (0.46)	0.84 (0.37)	0.31 (0.47)
MLD children	0.98 (0.16)	0.90 (0.32)	0.00 (0.00)	0.05 (0.22)	0.80 (0.41)	0.43 (0.50)
No "ToM" autistics	0.92 (0.28)	0.78 (0.42)	0.03 (0.17)	0.17 (0.38)	0.86 (0.35)	0.36 (0.49)
1st order autistics	1.00 (0.00)	0.78 (0.42)	0.11 (0.32)	0.22 (0.42)	0.94 (0.23)	0.44 (0.50)
2nd order autistics	1.00 (0.00)	0.95 (0.22)	0.00 (0.00)	0.05 (0.22)	0.86 (0.35)	0.36 (0.49)

It can be seen from table 3.2 that the performance of the subjects from the different groups was very similar. All manifest a moral/ conventional distinction in their judgements. Figure 3.4 depicts the influence of modifying the rule conditions (i.e., whether or not the teacher says that the act can be performed) on the children's judgements of whether moral and conventional transgressions are OK. From this it can be seen that all the groups of children in this study made a distinction between moral and conventional transgressions.

As stated in chapter one, the *permissibility* and the *seriousness* criteria both elicit the subject's judgements of the item's seriousness. They elicit value judgements. The *authority jurisdiction* criteria is a modifiability judgement. To be performed the child must calculate a new context for the transgression and then judge the transgression. Modifiability criterion judgements should be more precise measures of the child's moral/ conventional judgement of the transgression. For this reason, the *authority jurisdiction* judgements of each child were reanalysed separately. Table 3.3 reveals how many of the subjects in each of the five groups judged how many of the moral transgressions to be *independent of authority* and how many of the conventional transgressions to be *dependent on authority*.

Figure 3.4: Permissibility as a function of rule conditions.

Figure 3.4(a): Permissibility of moral and conventional transgressions under normal rule conditions (the *permissibility* question).

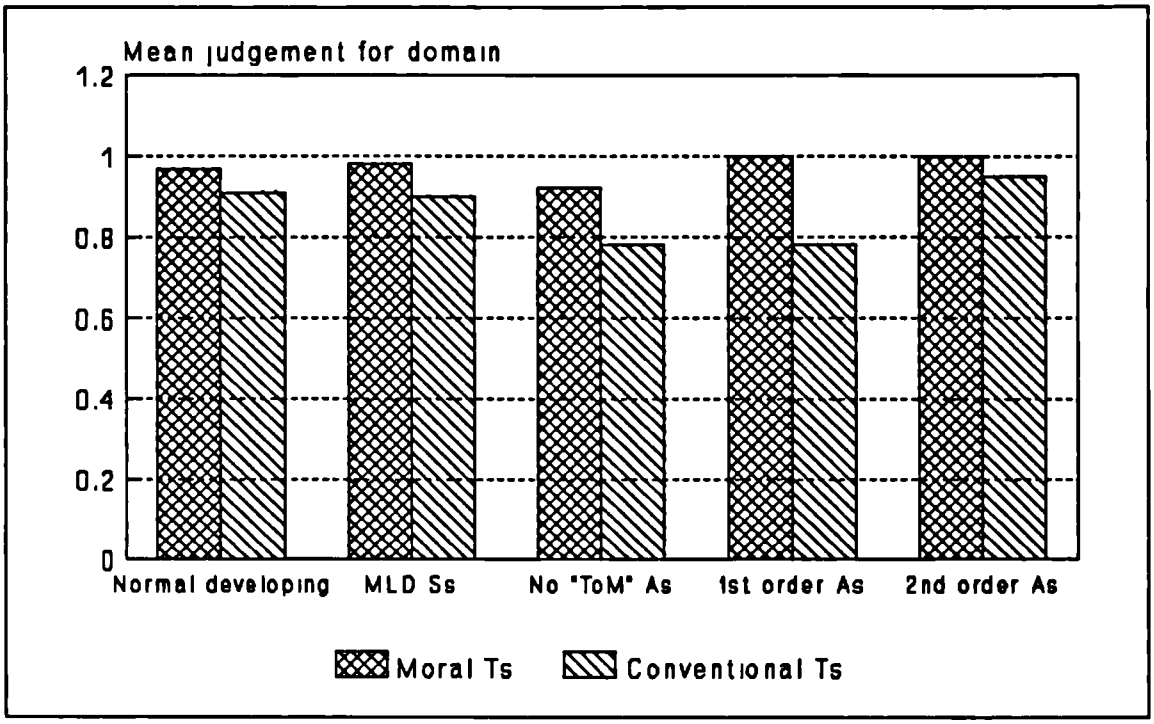


Figure 3.4(b): Permissibility of moral and conventional transgressions under modified rule conditions; the teacher having said that at this school anyone can do the act (the *authority jurisdiction* question).

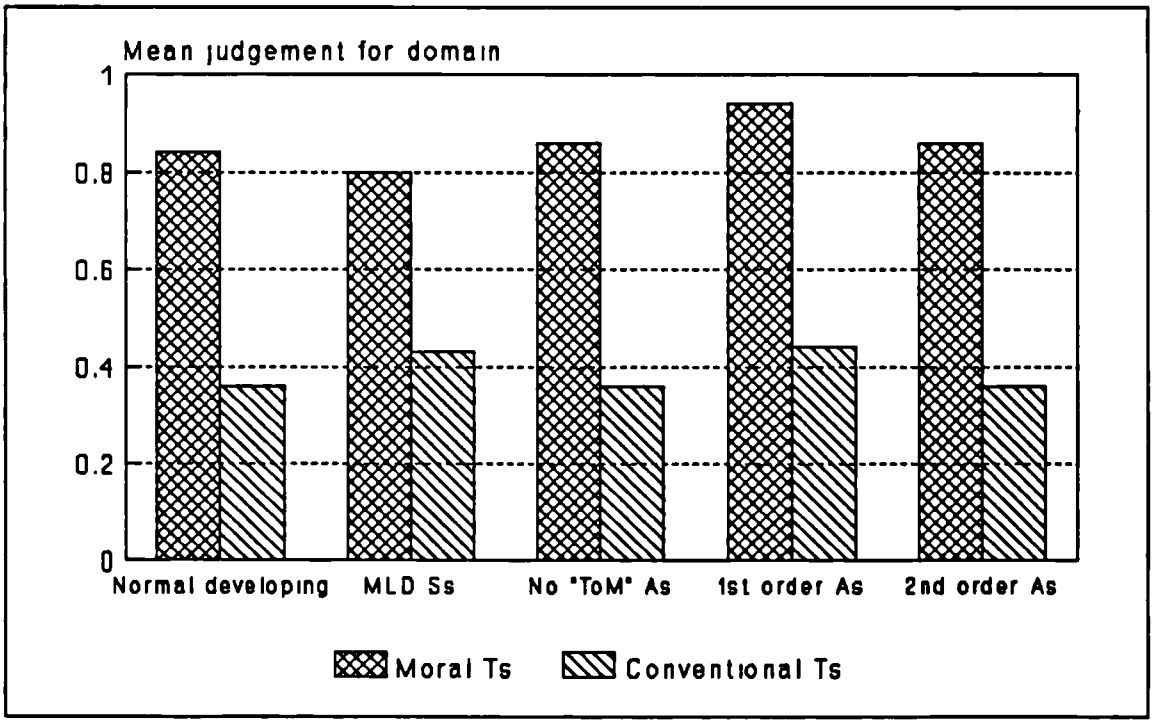


Table 3.3: Showing how many of the subjects in each of the five groups judged how many of the moral transgressions to be *independent of authority* and how many of the conventional transgressions to be *dependent on authority*.

	Moral			Conventional		
	2 + independent	1 independent	None independent	None dependent	1 dependent	All dependent
Normally developing	10	0	0	1	0	9
MLD	8	2	0	2	1	7
No "ToM" autistics	10	0	0	0	3	7
1st order autistics	10	0	0	1	3	6
2nd order autistics	5	0	0	0	0	5

Table 3.3 shows that all the subjects, apart from two MLD subjects, generally judged the moral transgressions to be *authority independent.* Table 3.3 shows that the subjects were much less likely to judge the conventional transgressions as *authority dependent* than they were to judge the moral transgressions as *authority independent.* However, most subjects in all five groups were generally likely to judge the conventional transgressions as *authority dependent.* There are no significant group differences.

3.4: Discussion

This experiment demonstrates firstly that most autistic children make a distinction between moral and conventional transgressions in their judgements. Secondly, the study shows that level of ability on "Theory of Mind" tasks is not associated with the tendency to distinguish moral and conventional transgressions. Thirdly, this experiment demonstrates that low IQ (the MLD and the autistic subjects who failed all the false belief tasks) does not prevent the child from making the moral/ conventional distinction. Returning to the DCM expressed in figure 3.2 it can be seen that this model is incorrect. ToMM, and thus role taking and empathy, are not prerequisites for the moral/ conventional distinction.

The importance of this finding must now be considered. Kohlberg (e.g., Kohlberg, 1981), Piaget (Piaget, 1932) and Turiel (Turiel, 1977) have all argued that role taking was implicated in the

development of morality. Piaget argued that role taking was one of the factors necessary for the formation of autonomous morality. Kohlberg suggested that an ability to role take was necessary if the child was to reach a post-conventional level of morality. Turiel (1977) claimed that role taking was one of the methods available to the child for the construction of moral and conventional concepts. While the data presented in this paper does not bear on Kohlberg and Piaget's claims (neither of these accounts would predict the existence of the moral/ conventional distinction) it is clear that role taking is not necessary for the formation of the moral/ conventional distinction.

As stated in chapter one, Turiel (1977) claimed that the individual has a variety of methods available to him for this construction of concepts; observation, communication, imitation and role-taking. As argued in chapter one, all of these methods (i.e., observation, communication and imitation), apart from role taking, are methods by which information is copied from one individual to another. Role taking is the only one of these methods that actually involves the construction of information by the child. Thus, if Turiel (1977) is to claim that the moral/ conventional distinction is not a direct product of cultural transmission (that the culture does not "impose" the domain structure on the child), he must consider that role taking is the principal method for the construction of the moral/ conventional distinction. As this experiment shows, if he does this he is wrong.

However, Turiel (Turiel, 1983) has described two other methods that result in "judgements of moral necessity" (Turiel, 1983; p. 43): personal experience of pain as a consequence of being the victim of another's moral transgression and; counterfactual reasoning (imagining whether it would have been more desirable for the transgression not to have been committed). Autistic children can presumably experience pain [though see Frith and Happé (1992) for discussion on whether they can reflect on this experience] and reason counterfactually. Can these two abilities compensate for the autistic child's inability to role take and allow for the development of the moral/ conventional distinction? There is no reason to believe that they can. An individual may have been the victim of a moral transgression. He/ she may have experienced pain. However, these experiences, in themselves, cannot result in a "judgement of moral necessity". There is no structure in the cognitive system that Turiel and his colleagues have described (Turiel, 1983; Turiel & Davidson, 1986; Turiel & Smetana, 1984) that would result in any motivation to reduce pain in *other* individuals. Similarly, there is no structure in the cognitive system that Turiel and his colleagues have described that would result in the belief that it would be more "desirable"

if a moral transgression were not committed. There is no reason, given the cognitive system Turiel has described, that an individual should consider a situation where another individual is hurt as less desirable than one in which they are not hurt. Whatever the structure that does result in this consideration is additional to the system Turiel has so far described.

Kagan (1984) claimed that empathic arousal to moral transgressions resulted in the moral/ conventional distinction (see chapter one). This study showed that the empathic reaction was not a prerequisite for the moral/ conventional distinction. However, Kagan also mentioned other emotional reactions to moral transgressions that would result in the moral/ conventional distinction. Kagan might argue that these other emotional reactions could compensate for the lack of the empathic reaction. Thus, this study cannot disprove Kagan's position. This study can only show that empathy is not a prerequisite for the distinction.

Of the forty-five subjects, one normally developing subject, two MLD subjects and one 1st order autistic subject failed to make the moral/ conventional distinction. These were all the subjects who judged all of the conventional transgressions to be *authority independent.* Other subjects made the distinction poorly but they were more likely to judge moral transgressions, than conventional transgressions, as *authority independent.* The question must be asked: Why did these subjects fail to make the distinction? The answer probably lies in the specifics of the task. A successful distinction necessitates that the subject be allowed to predispose the moral transgressions to one set of answers, the conventional transgressions to another. The child may over-rule VIM, the mechanism for this predisposition. Indeed, one of the two MLD children said "You're trying to trick me." It seems probable that these children were confused by the specifics of the task. There is certainly no reason yet to believe that these children lacked VIM.

It might be wondered what role ToMM plays in the development of morality. However, this question will only be truly answered here in relation to the moral/ conventional distinction. Presumably, ToMM, or at least a "Theory of Mind" is involved in the formation of the censure against lying and in concepts of guilt. As Morton (1987; p. 36) states: "To lie, one not only must believe that others have counter-factual beliefs, but one must believe that one can induce such beliefs." Without ToMM, an individual could not comprehend a lie. So without ToMM, an individual would not censure lying. Similarly, without ToMM the individual could not understand guilt. Guilt requires the intention to have committed the act (Hoffman, 1987). Thus, to understand guilt the individual must be able to understand intention. As regards the moral/

conventional distinction, ToMM may have a role in the formation of the justification categories (the reasons people give for why the act is wrong). ToMM may be a prerequisite for representing the emotional experience of other people. If autistic children cannot represent these vicarious experiences then they will not be able to state, unless they are instructed, why moral transgressions are wrong (i.e, because they hurt people). While VIM may operate and result in these individuals judging moral transgressions as bad, they will not be able to represent the experience of the victims and thus incorporate it as meta-knowledge as to why the act is bad.

Figure 3.5: A DCM of VIM incorporating the role of ToMM. The model shows ToMM as a prerequisite for *internally generated moral justifications* and *empathy*.

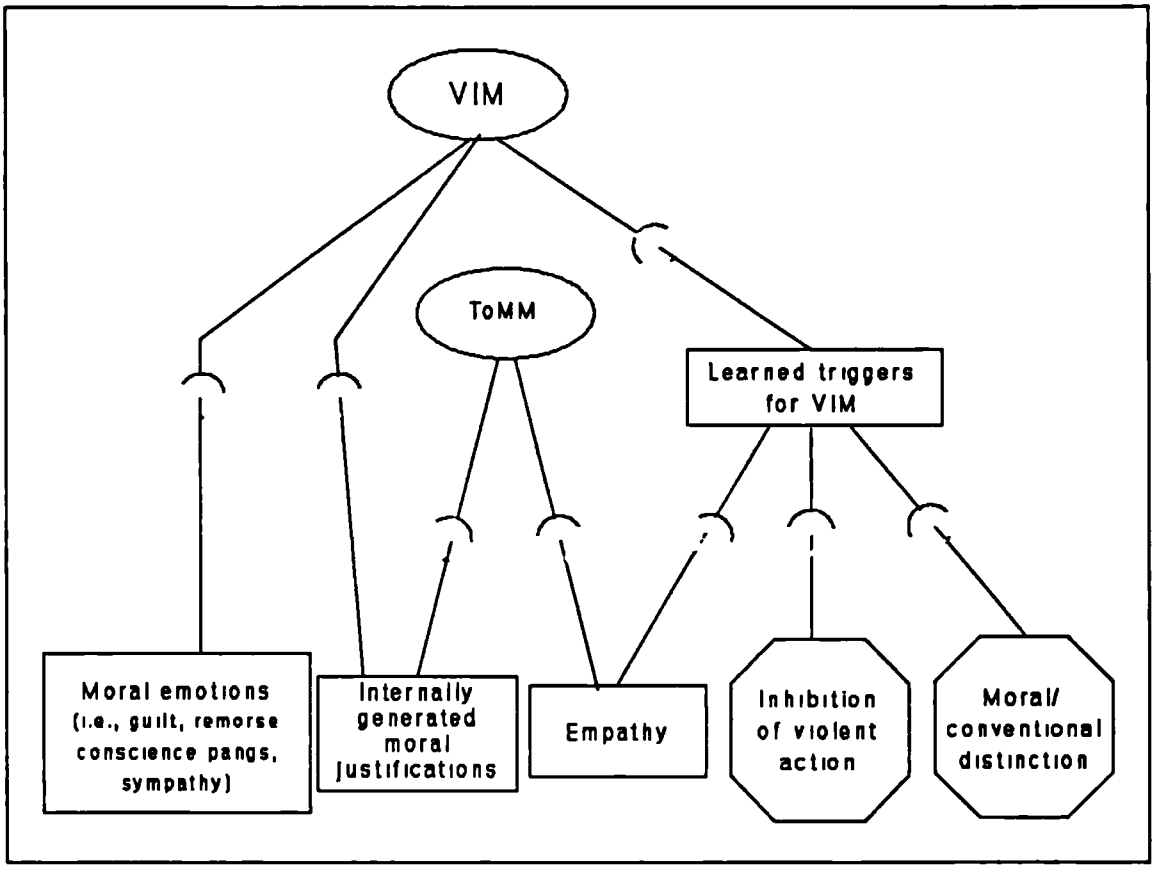


Figure 3.5 shows a DCM of VIM which incorporates ToMM. I show ToMM as a prerequisite for "internally generated moral justifications".[5] As stated above, without ToMM the individual

[5] This is not to say that an individual without ToMM (or VIM) will fail to make any moral justifications. An individually may be taught moral justifications; e.g., "moral rules are bad because they hurt people". I am only claiming that an individual without ToMM (or VIM) will fail to internally generate moral justifications.

cannot represent the experience of the victim. If the individual lacks representations of the experience of the victim, the individual will not be able to incorporate these representations into his theories of why moral transgressions are wrong; i.e., the individual will not be able to form internally generated moral justifications. Of course, if the individual lacks VIM, he will only judge moral transgressions as "bad" in that they involve the breaking of rules. Again, his meta-knowledge should stress rules and lack reference to victims.

In figure 3.5, I also show ToMM as a prerequisite for empathy. As stated above (see also chapter two; section 2.5.2), empathy is assumed to be a product of representations formed through role-taking. In this chapter, I have argued that ToMM is a prerequisite for role-taking to occur. Thus, ToMM must be a prerequisite for empathy.

3.5: Conclusions

This study has allowed several conclusions as to which cognitive functions are not prerequisites for the moral/ conventional distinction. This study reveals that a "Theory of Mind", role-taking and empathy are not prerequisites for the development of the distinction. Chapter four will investigate the moral/ conventional distinction in a subject population who might be expected to be lacking in VIM; Antisocial Personality Disorder subjects.

Chapter Four

Antisocial Personality Disorder And Moral Development

4.1: Introduction

In chapter two, I presented a Developmental Contingency Model (DCM) which showed the developmental consequences of the proposed Violence Inhibition Mechanism (VIM). In chapter three, I contrasted this model with a DCM which was based on Turiel's claim of the importance of role-taking in the development of the moral/ conventional distinction. In addition, I demonstrated that the Theory of Mind Mechanism (ToMM), and thus role-taking, were not prerequisites for the formation of the moral/ conventional distinction. Finally, I expanded the original DCM presented in chapter two to include the role of ToMM.

In the DCM (see figure 3.5), I proposed that VIM was a prerequisite for: the moral emotions (e.g., guilt, remorse and empathy); the inhibition of violent action; empathy and; the moral/ conventional distinction. Consequently, it can be predicted that an individual without VIM would show deficits in three areas: the experience of the moral emotions; the inhibition of violent action and; in making the moral/ conventional distinction (see figure 4.1 for a causal model showing the developmental consequences of a lack of VIM). This chapter will begin to explore whether Antisocial Personality Disorder may describe such an individual.

4.1.1: The concept of antisocial personality disorder

Pichot (1978) has traced the intellectual roots of antisocial personality disorder (APD) to Pritchard (1837). Pritchard developed the concept of "moral insanity" to account for socially damaging or irresponsible behaviour which was not associated with known forms of mental disorder. He attributed morally objectionable behaviour to be a consequence of a diseased "moral faculty". Pritchard's concept, according to Pichot, can be seen in the description of "moral imbeciles" in the English Mental Deficiency Act of 1913. This Act described "moral imbeciles" as "persons who from an early age display some permanent moral defect coupled with strong vicious or criminal propensities". This same concept, according to Pichot, is also

present in the English 1959 Mental Health Act as Psychopathic[1] Disorder. This act, and the revised Mental Health Act of 1983 define Psychopathic Disorder as, effectively, a "persistent disorder or disability of mind which results in abnormally aggressive or seriously irresponsible conduct and requires or is susceptible to medical treatment".

This definition is strictly a legal one; it has not proved to be clinically valid. As the Butler Committee (Home Office/ Department of Mental Health and Social Security, 1975) concluded: "The class of persons to whom the term 'psychopathic disorder' relates is not a single category identifiable by any medical, biological or psychological criteria". However, clinical definitions are available.

The concept of "sociopath" (Partridge, 1930) was incorporated into the first edition of DSM (American Psychiatric Association, 1952) as "sociopathic personality disturbance". The term sociopath was used in America in preference to psychopath to stress the revised assumption that the origins of the disorder were social (Blackburn, 1988a). However, this classification was a generic category which included antisocial reaction, "dyssocial reaction", sexual deviation, and addictions. This was obviously unsatisfactory and by the time of DSM-II (American Psychiatric Association, 1968) the term had been dropped. DSM-II included antisocial personality as a distinct category within the personality disorders. This category has been retained in DSM-III-R (American Psychiatric Association, 1987).

Within DSM-III-R, traits are central to the concept of personality. DSM-III-R defines traits as "enduring patterns of perceiving, relating to, and thinking about the environment and oneself". However, the DSM criteria for antisocial personality are predominantly antisocial types of *behaviour*, the only traits referred to are irritability and aggressiveness, impulsivity and recklessness, and none of these are essential to the diagnosis. This is inconsistent with the avowed aim of DSM to describe personality disorder in terms of traits. Indeed, Millon (1983) claims that this category is "an accusatory judgement rather than a dispassionate clinical formulation".

Others have attempted clinical descriptions of APD. For example, Cleckley developed a "clinical profile" of the psychopath from his extensive clinical experience with the disorder. Cleckley listed (and elaborated on) 16 key characteristics. These are: (1) superficial charm; (2)

---

[1] Psychopathic, as an adjective, was first used in mid-19th century Germany (Blackburn, 1988b). Originally, its definition was confined to its etymological meaning; "psychologically damaged". However, as can be seen, its definition came to be "morally impaired".

absence of psychotic signs; (3) absence of nervousness; (4) unreliability; (5) untruthfulness and insincerity; (6) lack of remorse or shame; (7) inadequately motivated antisocial behaviour; (8) failure to learn from experience; (9) egocentricity and incapacity for love; (10) emotional poverty; (11) lack of insight; (12) unresponsiveness in interpersonal relations; (13) uninviting behaviour, sometimes with alcohol; (14) empty suicide threats; (15) impersonal sex life; (16) failure to follow a life plan.

Cleckley's criteria correlate significantly with the DSM-III diagnosis of antisocial personality disorder (Hare, 1983; 1985a); diagnoses by Cleckley's criteria and DSM-III frequently concur. Unfortunately, however, Cleckley's criteria are not easy to use; they are time consuming and difficult to score. Because of this, Hare (1980; 1985b) developed an empirically based checklist (The Psychopathy Checklist [PCL]), which includes several of Cleckley's criteria, but which also includes items of social deviance, such as delinquency, promiscuous sexual relations, and "many types of offense". This measure also correlates significantly with the DSM-III diagnosis of antisocial personality (Hare, 1983, 1985a).

Several researchers have questioned the validity of APD as a concept (Blackburn, 1988b; Vaillant, 1975). For example, Blackburn has claimed that the concept is "little more than a moral judgement masquerading as a clinical diagnosis" (Blackburn, 1988b; p. 511). Their argument has centred on the claim that those classified as APD are not a homogeneous group. Blackburn points out that factor analysis of the checklist indicates that factors of personal deviance (impulsive life-style, egocentricity) are distinguishable from factors of social deviance (inappropriate sexual and parenting behaviour, antisocial history, and inadequately motivated criminal acts; see Hare, 1980; Raine, 1985).

Blackburn (1988b) argues from the presence of multiple factors implies that APD "is not a meaningful focus for theory and research"; that it does not describe a homogeneous group of individuals. However, it is questionable whether APD subjects should be expected to form a homogeneous group. As stated in chapter three, the presenting symptoms of autistics can be diverse. Any single cognitive deficit may have different behavioral implications for different individuals according to the relative presence/ absence of other, compensatory mechanisms. Homogeneity is only important at the cognitive level. It is intended that in this chapter I will examine whether there might be a cognitive deficit specific to APD.

4.1.2: Antisocial Personality Disorder and VIM

Figure 4.1 represents a causal model showing the developmental consequences of a lack of VIM. Causal models are divided into three levels: physiological, cognitive and behavioral (*see* Morton & Frith, *in press*). The relationship of connected elements within a causal model is one of causality. Thus, the model represents my claim that a lack of VIM causes an absence of the moral emotions. Elements within a causal model that are unaffected by the absence of another element are shown "protected" within boxes. Thus, in figure 4.1, the absence of VIM is shown to have no effect on the emergence of all other emotions.

Figure 4.1: A causal model of the developmental consequences of an absence of VIM.

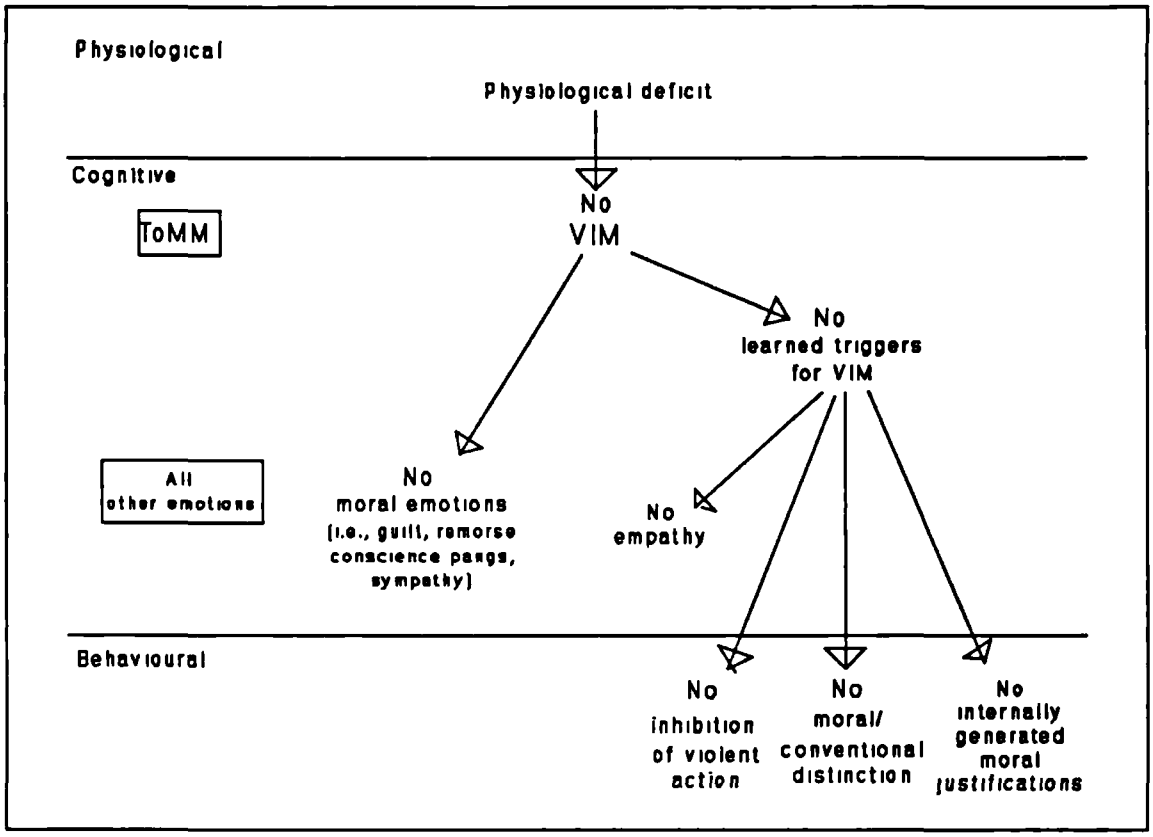


Figure 4.1 shows my claim that the absence of VIM is due to an underlying physiological deficit. It is suggested that some deficit at the physiological level results, at the cognitive level, in a lack of VIM. Obviously, an absence of VIM means that additional triggers will not be learnt for VIM. Thus, representations formed through role-taking will not activate VIM. Within the individual, this will result in an absence of empathy. As stated in chapter two, the activation of VIM results, through inter-schema competition, in the generation of arousal which through "meaning analysis" should be experienced as one of the moral emotions. Of course, and as

shown in figure 4.1, without VIM there will be no arousal generated in distress cue situations and thus no moral emotions. However, an absence of VIM will have no effect on the generation of arousal following other instances of inter-schema competition. Thus, I show in figure 4.1 all other emotions "protected" from an absence of VIM.

At the behavioral level, the predicted consequences of a lack of VIM are an absence of the inhibition of violent action and an absence of the moral/ conventional distinction. By suggesting that a lack of VIM will result in the lack of the inhibition of violent action, I do not mean to suggest that the individual without VIM will be incapable of inhibiting violent action. As stated in chapter two, there are at least two ways violent action can be inhibited on-line. One of these methods does not involve VIM; the SAS could deactivate the schema initiating the aggression. What I am implying here is that the absence of VIM will *developmentally* result in an absence of the inhibition of violent action. The child without VIM will not have their violent attacks negatively reinforced by distress cues. Thus, there will be no reduction in the probability of violence initiation over time. Indeed, if the violent attacks are reinforced through financial gain/ prestige there may be an increase in the frequency of attacks over time. The second behavioral predicted consequence of an absence of VIM is an absence of the moral/ conventional distinction. It is this prediction which will be tested in this chapter.

In section 4.1.1, I described several definitions of Antisocial Personality Disorder (APD). Some of these definitions of APD have stressed the absence of remorse, guilt and empathy (e.g., Cleckley, 1964; DSM-III-R, 1987; Hare, 1980; 1985b). As stated above, an absence of the moral emotions and an absence of empathy are the predicted developmental consequences of an absence of VIM (see figure 4.1). There is thus the suggestion that APD may be due to an absence of VIM.

One of the better clinical descriptions of APD and the one which is now used as a basis for all research in the area, was provided by Hare. This description is expressed as a Psychopathy Checklist (PCL) which consists of a list of 20 (Hare, 1985b) or 22 items (Hare, 1980) that are associated with APD. These items (shown in table 4.1) were found to best discriminate between prison inmates with low and high levels of psychopathy according to Cleckley's (1964) conception. In figure 4.2, I show in a second causal model how many of these items might be developmental consequences of an absence of VIM. Please note that these behaviours only *might* be developmental consequences of an absence of VIM. There may be, and I would claim that there are, compensatory social factors (e.g., social class, religion) that would ameliorate the effect of an absence of VIM. It is thus theoretically possible that there are individuals without

VIM who are not diagnosed as APD and may, indeed, never have committed a crime. In addition, it is theoretically possible that some of those who fulfil the criteria for APD on the PCL do not lack VIM. Other factors may have predisposed them to extreme crime. It is with these caveats in mind that figure 4.2 should be considered.

Table 4.1: Items on Hare's (1985b) Psychopathy Checklist[2].

(1)	Glibness/ superficial charm.
(2)	Grandiose sense of self-worth.
(3)	Need for stimulation/ proneness to boredom.
(4)	Pathological lying.
(5)	Conning/ manipulative.
(6)	Lack of remorse or guilt.
(7)	Shallow affect.
(8)	Callous/ lack of empathy.
(9)	Parasitic lifestyle.
(19)	Poor behavioral control.
(11)	Promiscuous sexual behaviour.
(12)	Early behavioral problems.
(13)	Lack of realistic, long-term goals.
(14)	Impulsivity.
(15)	Irresponsibility.
(16)	Failure to accept responsibility for own actions.
(17)	Many short term marital relationships.
(18)	Juvenile delinquency.
(19)	Revocation of conditional release.
(20)	Criminal versatility.

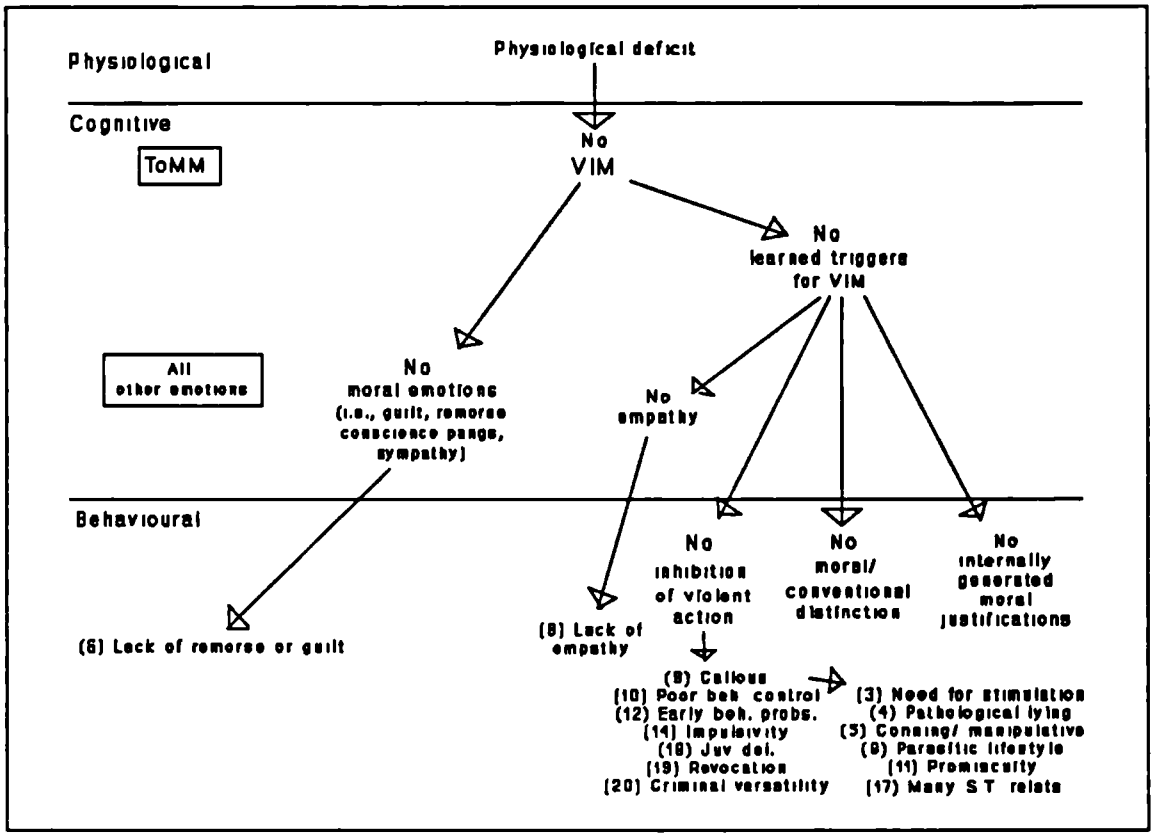
In figure 4.2, I show Hare's items at the behavioral level since they are identified through analysis of the subject's behaviour. For example, I show item 6 ("lack of remorse or guilt") as a behavioral consequence of the lack of moral emotions at the cognitive level. I also show item 8 ("lack of empathy") an a behavioral consequence of no empathy at the cognitive level.

In figure 4.1, I suggested that a developmental consequence of a lack of VIM was a lack of the developmental inhibition of violence. In figure 4.2, I suggest that this lack of a developmental inhibition of violence might give rise to item 8 ("callous"), item 10 ("poor behavioral control"), item 12 ("early behavioral problems"), item 14 ("impulsivity"), item 18 ("juvenile delinquency"), item 19 ("revocation of conditional release") and item 20 ("criminal versatility"). All of these

[2] All of Hare's items are terms for particular behavioral clusters. Thus, for example, item 8 ("callous") refers to a tendency, identified through the police records, to extreme violence.

items may reflect the increased probability of offense that a developmental lack of VIM would predict.

Figure 4.2: A causal model of the developmental consequences of an absence of VIM with reference to Hare's description of APD.

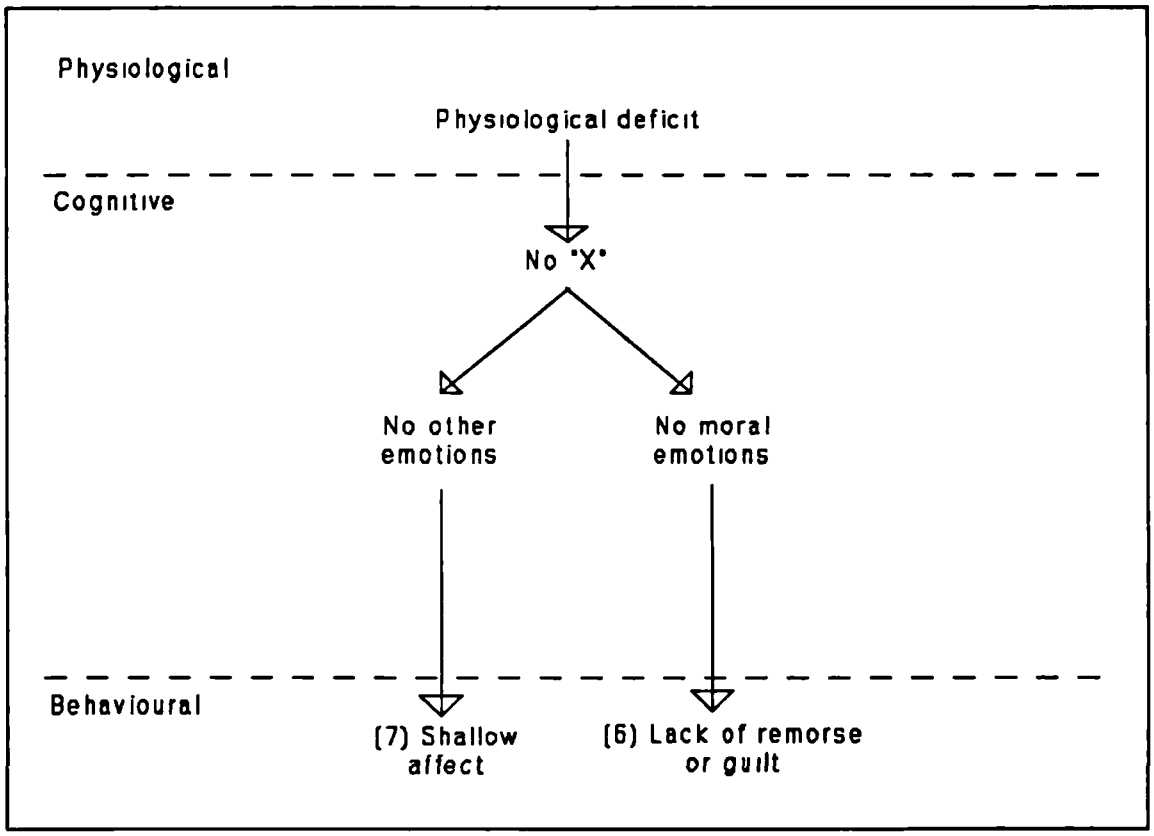


In figure 4.2, I have suggested that the items of the PCL should be distinguished into two categories; primary and secondary. All of the items I have discussed above would be classed as primary. They are predicted consequences of deficits at the cognitive level. However, some of the PCL items may be secondary; i.e., consequences of the behavioral consequences of a lack of VIM. In figure 4.2, I have shown item 3 ("need for stimulation/ proneness to boredom"; scored by reference to frequent job changes and drug/ alcohol use), item 4 ("pathological lying"), item 5 ("conning/ manipulative"), item 9 ("parasitic lifestyle"), item 11 ("promiscuous sexual behaviour") and item 17 ("many short term marital relationships") as secondary. These items would thus be predicted in individuals who possessed VIM but were involved in a life of crime. A life of crime might necessitate considerable lying (4), manipulation (5) and a "parasitic lifestyle" (9; the crime provides the income). A life of crime might allow drug use (3) and promiscuity (11). These behavioral problems might result in individuals without VIM being extremely difficult to live with. This might explain item 17.

Six other items have not been shown in this causal model: item 1 ("glibness/ superficial charm), item 2 ("grandiose sense of self-worth"), item 7 ("shallow affect"), item 13 ("lack of realistic, long term goals), item 15 ("irresponsibility") and item 16 ("failure to accept responsibility for own actions"). Accounts based on a VIM deficit might be provided for item 1, 2, 15 and 16 but these would be highly speculative. Without more data such accounts would not seem to be profitable. The VIM deficit hypothesis cannot explain item 7 or 13.

Of course, the model presented in figure 4.2 does not imply that the absence of guilt necessitates that the individual must also lack VIM. Other factors may be prerequisites for the development of the moral emotions and, as such, in their absence cause the lack of the moral emotions. One of the items the VIM account of APD cannot explain is the general shallowness of affect associated with APD. It is possible that there is some factor ("X") that is a prerequisite for the development of all emotions. The lack of this factor would cause the absence of all emotions. Figure 4.3 depicts a causal model of this alternative position. In chapter 7, this alternative position is tested.

Figure 4.3: A causal model showing item 6 ("lack of remorse or guilt" and item 7 ("shallow affect") as a consequence of an unspecified deficit responsible for the development of all emotions.



The causal model shown in figure 4.1 and 4.2 makes one prediction that is independent of the

clinical description of APD. This prediction is that an individual without VIM will lack the moral/ conventional distinction. It is thus predicted that if APD is due to an absence of VIM, APD subjects will fail to make the moral/ conventional distinction. It is this prediction that will be tested in this chapter.

Several previous studies have investigated the moral representations of the APD. However, this work has been conducted within the paradigmatic framework of Kohlberg (*see* chapter one). While it appears clear that the moral reasoning of delinquents is at a lower level than that of normal controls[3] (*see* Blasi, 1980; Trevathan & Walker, 1989) it is more debatable whether the moral reasoning of APDs is at a lower level than that of criminal controls. Fodor (1973) found that the moral reasoning of psychopathic youths was at a lower level than the moral reasoning of other delinquents. Campagna and Harter (1975) found that the moral reasoning of sociopaths to be lower than that of non-incarcerated normals, even when controlling for mental age. Jurkovic and Prentice (1977) found that psychopaths give evidence of less mature moral reasoning than other groups of delinquent and normal youths. However, Lee and Prentice (1988) only found that delinquents responded at a lower level than non-delinquents; the psychopaths did not reason at a lower level than the other delinquent groups.

Also, the above studies used scales of psychopathy (e.g., Quay's Behaviour Problem Checklist) that are of doubtful validity (see Hare & Cox, 1978). This methodological deficit was remedied by Trevathan and Walker (1989) who utilized Hare's Psychopathy Checklist. Trevathan and Walker (1989) observed a tendency for the psychopaths to reason at a lower level than non-psychopathic controls but this was not significant. However, both groups of delinquents scored at a significantly lower level than non-incarcerated controls. Thus, while it is clear that criminal groups may reason at a lower level than non-criminal controls, it is uncertain whether the moral reasoning of APD subjects is lower than that of other criminal groups.

No previous work has investigated whether the APD makes a distinction between moral and conventional rules. This is despite the fact that this is a fundamental aspect of the normal development of morality. If APD show a lack of this distinction it indicates that the failure in APD is very basic and has occurred at an early age. This chapter will describe an investigation into the moral/ conventional distinction of APD and non-antisocial incarcerated control subjects.

As stated in chapter one, the moral/ conventional distinction is manifested in both the individuals

---

[3] Though the reasons for this are not (see Blasi, 1980).

criterion judgements (i.e., *rule contingency*; "Would it [*the transgression*] be OK if there was no rule about it here?") and in his justification categories (the reasons the subject gives, for example, for judging the act not permissable). Up till now, I have concentrated on the individual's criterion judgements. In chapter one, I mentioned the types of *meta-knowledge* associated with moral transgressions; e.g., references to the *victim's welfare, appeals to fairness.* In chapter three, I mentioned that ToMM may be a prerequisite for the internal generation[4] of the above two types of moral meta-knowledge. Without ToMM, the individual cannot represent how another feels and therefore he cannot use this information in the formation of rules of action (i.e., in this case, for the formation of moral meta-knowledge). I suggest that if the subject does not know how another feels he is unlikely to make reference to how another feels when stating why he considers a moral transgression to be bad.

In this chapter, I suggest that VIM may also be a prerequisite for the internal generation of moral meta-knowledge; i.e., the tendency to make references to the *victim's welfare* and to *appeals to fairness.* In chapter two, I suggested that moral transgressions resulted in the activation of VIM which resulted in a withdrawal response which, through "meaning analysis" resulted in the transgression being judged bad. In this chapter I suggest that when asked why the act is bad, the individual should be able to do some sort of causal analysis which will determine that the distress to the other is the object which activated the withdrawal response; i.e., it is the object that is bad. I suggest that without VIM the individual will judge acts as bad only because he has been told that they are bad (by parents/ peers). Without VIM, if the subject is asked about why the act is bad, he will make reference to what he has been told.

Thus, the predicted results of this experiment were:

(1) That APD subjects will not make a distinction between moral and conventional rules.

(2) That APD subjects would treat moral rules as if they were conventional; that is, under permission conditions, the APD would say that moral as well as conventional transgressions were OK to do.

(3) That APD subjects would be less likely to make references to the pain or discomfort of victims than the criminal controls CCs.

---

[4] All forms of meta-knowledge could be learnt through direct cultural transmission. A parent telling a child that it is wrong to hit people because it hurts them would, assuming the child remembered the admonition, be an example of cultural transmission of moral meta-knowledge.

4.2: Method

4.2.1: Design

The experiment involved a 2 x 2 repeated measures factorial design. The independent variables were the 2 different subject groups (antisocial personality and non-antisocial incarcerated controls) and the two different domains of story (moral and conventional). The dependent variable was the responses of the subjects to the questions about the transgression situation.

4.2.2: Subjects

20 antisocial personality (N = 10) and criminal controls (N = 10) took part in this study. All were obtained through contacts in Broadmoor and Ashworth Special Hospitals and had been admitted to the Hospitals under the legal category of Psychopathic Disorder. The files of all the subjects were examined to obtain a PCL score in accordance with the guidelines of Hare (1985b), Wong (1985; cited by Hare, 1985b) having shown that PCL scores derived entirely from file data can be valid and reliable. Five items on the PCL were not scored: items 1, 2, 4, 13. These items were neglected because of the difficulty of obtaining such information from the files. However, Hare states that "as many as 5 items can be omitted without any appreciable reduction in reliability" (Hare, 1985b; p. 10). The subjects were then divided into two groups according to their score on the PCL; one high for psychopathy (the APD group), one low for psychopathy[5] (the criminal controls [CCs]). All subjects were male and white. Full subject characteristics are shown in table 4.2.

Two way ANOVAs (comparing the two subject groups) were undertaken for each of the subject criteria. These revealed no significant differences in age between the two groups ($F(1,19) = 1.83$; n.s.). The two tests of intelligence revealed no significant differences between the groups either when using the Wais ($F(1,17)$[6] $= 0.001$; n.s.) or Ravens matrices ($F(1,19) = 0.005$; n.s.). However, there was a significant difference between the groups on their psychopathy scores ($F(1,19) = 59.3$; $p < 0.001$).

---

[5] Please note that though these subjects scored lower on the PCL they were in Special Hospitals. All were murderers. Indeed, the only subject in the pool who had not killed was one of the APDs.

[6] Wais scores were obtained from the patients files and not by the present experimenter. Wais scores were not available for all subjects.

Table 4.2: Means for each of the subject criteria (standard deviations in brackets).

Group	N	Age	IQ: (Wais)	IQ: Ravens Matrices	Hare's Psychopathy score[7]
Antisocial personality	10	33.3 (7.7)	91.6 (17.2)	94.3 (30.5)	30.4 (2.1)
Criminal controls	10	37.8 (9.9)	91.8 (16.4)	92.1 (20.3)	15.2 (5.3)

In table 4.3, I show scores for the two groups broken down by item on the PCL. Separate ANOVAs were conducted for each of these items. The significance of the difference between the groups for each of these items is also shown in table 4.3. As can be seen, while APDs scored higher on all of the items, they only significantly did so on eight.

4.2.3: Materials

Ravens Matrices: This involves the subject selecting which of twelve tabs presented at the bottom of an A4 sized piece of paper would fit into the vacant slot in a pattern presented at the top of the same piece of paper. Twelve of these patterns were presented to each subject.

The moral/ conventional task: The stories that were used were identical to those described in chapter three.

Subject's responses to questions were recorded on standard scoring sheets.

[7] To gain a score out of 40 when only 16 items were being measured the score from the 16 items (out of 32) was multiplied by 40/32.

Table 4.3: APDs and CCs compared on individual items on the PCL.

	APD	CC	Significance
(1) Glibness/ superficial charm	--	--	--
(2) Grandiose sense of self-worth	--	--	--
(3) Need for stimulation/ proneness to boredom	1.66 (0.58)	0.89 (0.78)	0.06
(4) Pathological lying	--	--	--
(5) Conning/ manipulative	0.56 (0.88)	0.22 (0.67)	n.s.
(6) Lack of remorse or guilt	1.78 (0.44)	0.56 (0.88)	0.01
(7) Shallow affect	1.44 (0.88)	0.67 (0.71)	0.06
(8) Callous/ lack of empathy	1.94 (0.17)	1.22 (0.67)	0.01
(9) Parasitic lifestyle	0.82 (0.75)	0.67 (0.87)	n.s.
(10) Poor behavioral control	1.72 (0.44)	1.55 (0.53)	n.s.
(11) Promiscuous sexual behaviour	1.00 (0.98)	0.39 (0.78)	n.s.
(12) Early behavioral problems	2.00 (0.00)	0.78 (0.97)	0.01
(13) Lack of realistic, long term goals	--	--	--
(14) Impulsivity	1.94 (0.17)	1.22 (0.83)	0.05
(15) Irresponsibility	1.28 (0.96)	0.68 (0.74)	n.s.
(16) Failure to accept responsibility for own actions	0.54 (0.86)	0.44 (0.73)	n.s.
(17) Many short term marital relationships	0.77 (0.87)	0.22 (0.67)	n.s.
(18) Juvenile delinquency	1.89 (0.33)	0.56 (0.73)	0.001
(19) Revocation of conditional release	1.82 (0.67)	1.56 (0.88)	n.s.
(20) Criminal versatility	1.56 (0.53)	0.67 (0.50)	0.01

4.2.4: Procedure

Subjects were tested in one of the interview rooms attached to whichever ward the subject was housed on. Before the study commenced the subjects were introduced to the experimenter and informed about what they were to do. Subject consent forms were taken.

Before any of the transgressions scenes were read out to the subjects, they were informed that all of the scenes would occur within a school environment. It was decided to place the transgressions scenes within a school environment, as opposed to a ward or other adult environment, because piloting had shown that teachers were regarded by the subjects as legitimate authority figures for children. Some subjects did not regard nurses as legitimate authority figures for other adults.

Each of the transgression scenes was read out to the subject one at a time. The order of presentation of the transgression scenes was randomized across subjects. The transgression scenes were identical to those that have been described in chapter three. After the transgression scene had been presented, the subject was asked four questions:

(1) "Was it O.K. for X to do Y?" (ascertaining the subject's judgement of the *permissibility* of the act).

(2) "Was it bad for X to do [the transgression?]" and then "On a scale of one to ten, how bad was it for X to do [the transgression]?" (ascertaining the subject's judgement of the *seriousness* of the act).

(3) "Why was it bad for X to do [the transgression]?" (ascertaining the subject's *justification categories* for the act).

The subject was then told:

> "Now what if the teacher said before the lesson, before X did [the transgression], that "At this school anybody can Y if they want to. Anybody can Y."

and then asked a final question:

(4) "Would it be O.K. for X to do Y if the teacher says X can?" (assessing the rule's *authority jurisdiction*).

All responses were recorded by hand on a standard scoring sheet.

4.2.5: Scoring Procedure

The scoring procedure followed that commonly used in the literature (e.g., Smetana, 1981; Smetana & Braeges, 1990) and was identical to that used in chapter three. The answers to all questions, except 3, were scored categorically. *Yes* responses were assigned a score of 0, and *no* (*not OK*) responses a score of 1. Question 3 was scored according to the value (between 1 and 10) the subject had given that transgression. The justifications of the subjects were scored according to categories that had been used in previous research (Smetana, 1985). Two coders scored all justifications, and inter-rater reliability was high 91%.

4.3: Results

4.3.1: Criterion Judgements

Table 4.4 presents the means and standard deviations of moral and conventional judgements for each of the criterion judgements for both subject groups. Three 2(Domain) x 2(Group) ANOVAs were performed on the subject's responses for each of the three criterion judgements[8]. Because all responses were expressed as proportions, arcsine transformations were performed to correct for non-normality (Winer, 1971). These three ANOVAs revealed main effects of domain for all three judgements: *permissibility*, $F(1, 39) = 7.56$, $p = 0.05$; *seriousness*, $F(1, 39) = 8.53$, $p < 0.01$, and: *authority jurisdiction* (modifiability), $F(1, 39) = 24.57$, $p < 0.001$). Moral transgressions were judged significantly less *permissable*, more *serious* and less *authority dependent* than conventional transgressions.

Significant group differences were only shown in the results of the ANOVA on the *authority jurisdiction* (modifiability) criterion judgement. This ANOVA showed a main effect of group $F(1, 39) = 4.43$, $p = 0.05$ and a significant domain x group interaction $F(1, 39) = 8.18$, $p < 0.01$. However, a simple effects analysis using two way ANOVAs to examine the moral/ conventional

[8] Lunney (1970) has justified parametric analyses of variance (ANOVAs) with dichotomous data.

distinction of the two groups independently revealed different patterns of responding for the two groups. The CCs made a significant moral/ conventional distinction on all three criterion judgements (*permissibility* $F(1,19) = 8.92$, $p < 0.05$; *seriousness* $F(1,19) = 8.49$, $p < 0.05$; *authority jurisdiction* $F(1,19) = 43.69$, $p < 0.001$). However, the APD subjects did not make a significant moral/ conventional distinction on any of the criterion judgements (*permissibility* ($F(1,19) = 0.56$, $p <$ n.s.; *seriousness* $F(1,19) = 2.03$, $p =$ n.s.; *authority jurisdiction* $F(1,19) = 2.08$, $p =$ n.s.).

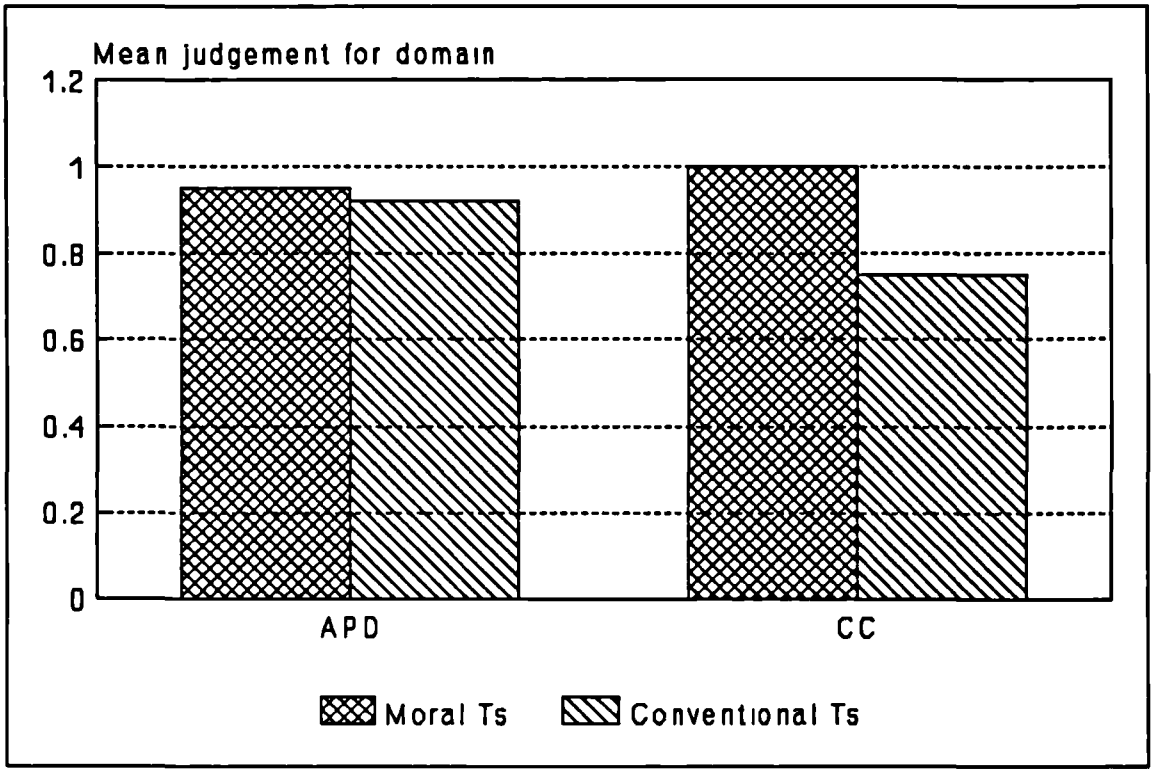
Table 4.4: The means and standard deviations of moral and conventional judgements for each of the criterion judgements for each of the subject groups.

	Criterion Judgements					
	Permissibility		Seriousness		Modifiability	
	M	C	M	C	M	C
APD	0.95 (0.16)	0.92 (0.27)	8.42 (2.33)	6.59 (3.66)	0.95 (0.22)	0.79 (0.41)
Criminal controls	1.00 (0.00)	0.75 (0.44)	8.06 (2.32)	4.47 (3.40)	1.00 (0.00)	0.42 (0.50)

It seems therefore that while the responding of the APD and CC subjects was only significantly different for the *authority jurisdiction* (modifiability) criterion judgement, the two groups can be differentiated (see figure 4.4). As predicted (prediction 1; see the simple effects analysis), the APD subjects did not show a moral/ conventional distinction on any of the criterion judgements. However, in contrast to predictions (prediction 2), the APD subjects did not judge moral transgressions as conventional on the *authority jurisdiction* criterion judgement; i.e., *authority dependent.* Indeed, APD subjects did the opposite, judging conventional transgressions as moral on this criterion judgement; i.e., not *authority independent.*

Analysis of individual subject data reveals the difference in the pattern of responding of the two subject groups even more clearly. Table 4.5 reveals the differences between individuals in the two groups in their pattern of responding on the *authority jurisdiction* question. Table 4.5 reveals how many of the subjects in each of the two groups judged how many of the conventional transgressions to be *authority independent.* All of the subjects (other than one APD subject) judged that all of the moral transgressions were *authority independent.*

Figure 4.4: Showing the results of the APD and CC subjects on the *permissibility* (figure 4.4a) and *modifiability* (figure 4.4b) questions.



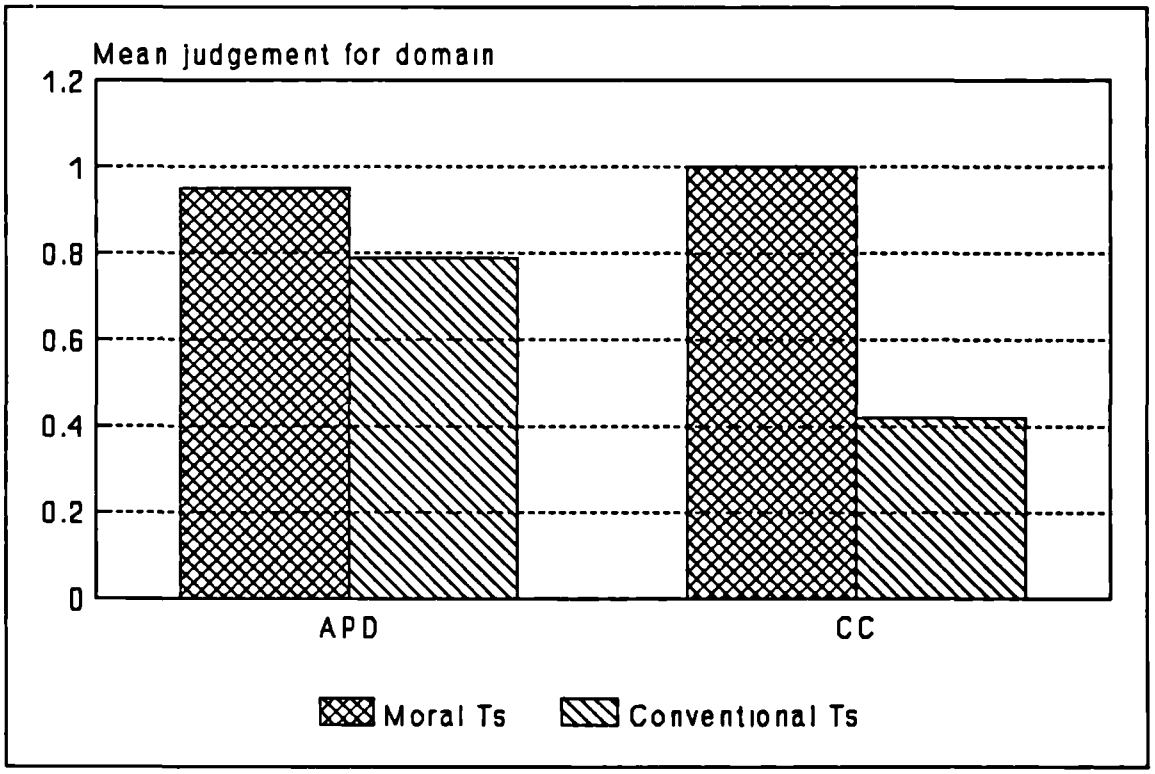


Table 4.5 clearly shows the difference between the two groups. 6 APD (as opposed to 1 CC) subjects did not distinguish between moral and conventional transgressions on the *authority jurisdiction* question at all; all of these subjects thought that the transgression was not OK even if the teacher said that it was. In addition, a two (Group) way ANOVA, performed on the "Quality of MvC Distinction" score, revealed that the APDs were judging significantly more of the conventional transgressions as moral ($F(1,19) = 8.10$; $p < 0.05$).

Table 4.5: The number of Antisocial Personality Disorder (APD) and Criminal Control (CC) in each of the "Quality of MvC Distinction" categories. No distinction = no transgressions were judged *authority independent*; mild distinction = 1 transgression was judged *authority independent*; clear distinction = 2 or more transgressions were judged *authority dependent*.

Group	Quality of MvC Distinction		
	No distinction	Mild distinction	Clear distinction
APD	6	2	2
CC	1	1	8

Only 2 APD (as opposed to 8 CC) subjects made a clear moral/ conventional transgression (i.e., considered that more than two of the conventional transgressions were OK to do under the permission conditions). Even then, 1 of these 2 APD subjects actually viewed all transgressions, apart from the 2 physical violence transgressions, as permissable under permission conditions; this subject, unlike all the other 19 studied, considered that property damage, under permission conditions, was permissable.

4.3.2: Justification Categories

Table 4.6 and figure 4.5 depict subjects' proportionate use of justifications for (combined) moral and conventional items. It seems that, regardless of group, *victim's welfare* reasoning was more commonly used to justify moral items while *disorder* statements and *rudeness* were more commonly used to justify conventional items. 2(group) x 2(domain) ANOVA[9]s, performed

[9] Because all responses were expressed as proportions, arcsine transformations were performed to correct for non-normality (Winer, 1971).

separately for each of the justification categories, confirmed this impression. Main effects for domain were found for *victim's welfare* ($F(1, 39) = 18.17$; $p < 0.001$) and *disorder statements* ($F(1, 39) = 7.14$; $p < 0.05$).

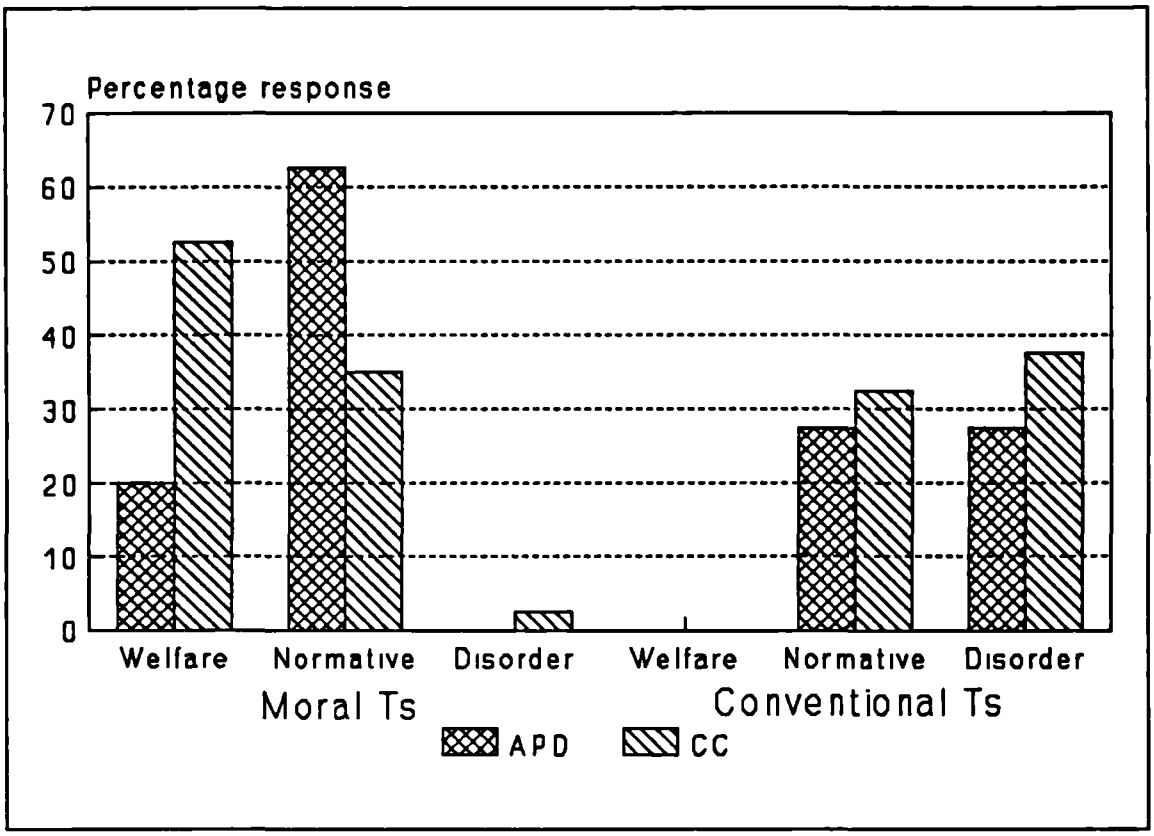
The ANOVAs also revealed a main effect of group ($F(1, 39) = 6.50$; $p = 0.05$) for *victim's welfare*. As predicted (prediction 3), APD subjects are significantly less likely to justify items by references to the *victim's welfare*. There was also a significant group by domain interaction ($F(1, 39) = 16.50$; $p = 0.05$). This was a product of the fact that this difference between the groups was only present for the moral items; no conventional items were justified through references to *victim's welfare*.

Examination of *victim's welfare* justification by the individual subject revealed that 5 of the APD and 9 of the CC subjects used this form of justification at least once. Of the 5 APD subjects who used this justification, 2 were the 2 subjects who showed a clear moral/ conventional distinction (see table 4.5) and 1 was 1 of the subjects who made a mild moral/ conventional distinction. The 1 CC subject who did not use *other's welfare* justifications did, however, make a clear moral/ conventional distinction.

Table 4.6: The proportionate use by the Antisocial Personality Disordered (APD) and Criminal Control (CC) subjects of the justification categories.

	Group			
	APD		CC	
	M	C	M	C
Other's welfare	20.00	0.00	52.50	0.00
Normative references	62.50	27.50	35.00	32.50
Disorder statements	0.00	27.50	2.50	37.50
Lack of change	12.50	7.50	2.50	12.50
Rudeness	0.00	27.50	0.00	10.00
Other	5.00	10.00	5.00	7.50

Figure 4.5: Showing the proportionate use of justification categories by APD and CC subjects for moral and conventional transgressions.



4.3.3: Individual item PCL scores and the moral/ conventional distinction

It was decided to investigate the relationship between the subject's score on each of the items of the PCL and his tendencies to judge conventional transgressions as moral and to make references to *victim's welfare* in his justifications. In table 4.7, I show those PCL items which significantly correlated with tendency to judge conventional transgressions as moral and tendency to make *victim's welfare* justifications. In table 4.8, I show the intercorrelations between total PCL score, tendency to judge conventional transgressions as moral and tendency to make *victim's welfare* justifications.

These two tables show only two significant correlations of individual PCL items with tendency to judge conventional transgressions as moral ("lack of remorse or guilt" and "criminal versatility"). Total PCL score does not correlate significantly with this tendency. However, five individual test items and total PCL score correlates significantly with the tendency to make *victim's welfare* justifications. In addition tendency to make *victim's welfare* justifications

correlates significantly with tendency to judge conventional transgressions as moral.

Table 4.7: Individual items on the PCL which significantly correlated with tendency to judge conventional transgressions as moral and tendency to make *victim's welfare* justifications.

PCL item	Tendency to judge conventional transgressions as moral	Tendency to make *victim's welfare* justifications
Lack of remorse or guilt	0.70**	-0.47
Callous/ lack of empathy	n.s.	- 0.54
Early behavioral problems	n.s.	- 0.59*
Juvenile delinquency	n.s.	- 0.56
Criminal versatility	0.50	- 0.49

(* = $p < 0.01$; ** = $p < 0.001$)

Table 4.8: Intercorrelations of total PCL score, tendency to judge conventional transgressions as moral and tendency to make *victim's welfare* justifications.

	Tendency to judge conventional transgressions as moral	Tendency to make *victim's welfare* justifications
Total PCL score	0.39[8]	- 0.53
Tendency to make *victim's welfare* justifications	- 0.45	

4.4: Discussion

The present study examined the form of the moral/ conventional distinction made by Antisocial Personality Disorder (APD) and Criminal Control (CC) subjects and the categories used by these subjects when they justify their judgements. This study revealed: Firstly, and in line with predictions, that while CC subjects made the moral/ conventional distinction, APD subjects did not; Secondly, and in contrast with predictions, that APD subjects treated conventional transgressions like moral transgressions rather than treating moral transgressions like conventional transgressions and; Thirdly, and in line with predictions, that APD subjects were much less likely to justify their items with reference to *victim's welfare.*

[8] This correlation is not significant.

It should be noted that while these results were broadly in line with predictions, they would not be expected from an analysis of the literature. As shown in chapter one, the observation in individuals of a moral/ conventional distinction is a particularly robust phenomenon; it is found across age (e.g., Nucci, 1981), across cultures (e.g., Song, Smetana & Kim, 1987) and across ability levels (e.g., chapter three, this volume). All other populations examined have been found to make this distinction.

It should also be noted that these findings cannot be explained as a result of poor parenting strategies (this includes neglect or child abuse). While clinicians have reported that many APD subjects have been abused as children, not all have, nor have the CC population used here been free of this abuse. More importantly, Smetana, Kelly and Twentyman (1984) examining the moral/ conventional distinction in abused children, found that these children did make the distinction.

Two of these findings were predicted by the model presented in chapter two and depicted as figure 4.1. This model suggested that VIM has developmental implications for moral reasoning. This model made the claim that VIM is a prerequisite for the development of the moral/ conventional distinction. This model also suggested that VIM was a prerequisite for the internal generation of moral meta-knowledge; i.e., *victim's welfare* and *appeals to fairness*. The third finding was opposite to my prediction.

I suggested in chapter two, that when an individual learns of a moral transgression, VIM is activated by the presence of a victim. I suggested that VIM activation results in an inhibitory response resulting in arousal. This arousal, associated with the withdrawal response, is experienced as aversive. The act which elicited this arousal is associated with the aversiveness; it is considered undesirable. Even if the transgression situation is changed, i.e., there is no rule prohibiting the moral transgression, VIM will still be activated by the presence of the victim, and so the act will still be considered not O.K. to do. In contrast, when an individual processes a conventional transgression VIM will not be activated and there will be no aversive arousal. The individual will therefore consider that any conventional transgression that is not prohibited by a rule is OK to do.

This account implies: Firstly, that the APD should judge moral and conventional transgressions similarly and; Secondly, that the APD should process all transgressions as conventional (*authority dependent*) because, given the lack of VIM, no aversive arousal should be generated. As stated above, the first prediction was confirmed; APD subjects did judge moral and

conventional transgressions similarly. However, APD subjects judged all transgressions as moral, not conventional. However, this second finding is not incompatible with the VIM position. Indeed, perhaps this finding is not surprising. These subjects were all incarcerated and presumably motivated to be released. All wished to demonstrate that the treatments they were receiving were effective. They therefore would be motivated to show how they had learned the rules of society (notice the predominance of normative statement justifications from both groups; table 4.6). The APDs manifest this desire on the *authority jurisdiction* criterion judgement, by suggesting that all transgressions are *authority independent.* I suggest that this is because the APDs lack VIM and thus are unable to identify the distinguishing features differentiating moral and conventional transgressions. This inability coupled with a desire to demonstrate adherence to societal rules results in their judgement of all the transgressions as *authority independent.* The CCs, in contrast, though presumably equally motivated to be released, are incapable of ignoring the distinguishing features of moral and conventional transgressions because of the operation of VIM, and thus answer the *authority jurisdiction* question appropriately.

Examining the justifications produced by subjects when they were explaining why they thought the transgression were not O.K. to do (see table 4.6 and figure 4.5), it can be seen that APD and CC subjects used similar justifications if the transgression was conventional but not if the transgression was moral. If the transgression was moral, CC subjects used predominantly *other's welfare* justifications ("it hurts") while APD subjects used predominantly *normative* justifications (i.e., "it's wrong" or "it's not socially acceptable"). This result was as predicted. Without VIM, the individual may not associate the pain of the other with the transgression and thus will not justify the acts wrongness by referring to the welfare of others.

It could be suggested, from table 4.6 and figure 4.5, that the APD subjects, though they failed to make a moral/ conventional distinction in their criterion judgements, are making a moral/ conventional distinction in their justifications. They certainly do show a tendency to give different justifications for moral and conventional transgressions. However, this cannot be taken as evidence against the position proposed here. In chapter one, I defined the moral/ conventional transgressions as the distinction in an individual's *criterion judgements* between moral and conventional transgressions. I suggested that the activation of VIM by representations of moral, but not conventional, transgressions was responsible for this distinction. I also suggested that VIM was a prerequisite for the internal generation of justifications centred on the plight of the victim. However, I did not suggest that VIM had a role in the generation of any other form of justification category. Indeed, I do not believe that it does. What justification category an

individual gives for a specific transgression will be a function of the salient aspects of that transgression. There is no reason to believe that the salient features of a moral transgression should be identical to those of a conventional transgression (outside of the fact that moral transgressions necessarily result in victims). Indeed, there is reason to believe that they should not be. Moral transgressions (e.g., one individual hitting another) need not result in classroom disorder. Conventional transgressions (e.g., talking in class) necessitate classroom disorder. Thus, the fact that the APD subject give different, non-victim based justifications for moral and conventional transgression cannot be used against the position being advocated here.

4.5: Conclusions

This study has produced confirmed two predictions of the causal model presented as figure 4.1. APD subjects are significantly more likely to fail to make the moral/ conventional distinction and they are significantly less likely to make reference to the welfare of others. While this study has not proven that APD subjects lack VIM, it has provided evidence that is in line with position. However, there are other explanations of APD. One of these is that APD subjects suffer from an inability to role take. This possibility will be tested in chapter five.

Chapter Five

Antisocial Personality Disorder and the "Theory of Mind"

5.1: Introduction

Chapter four revealed that the antisocial personality failed to make a moral/ conventional distinction and that the Antisocial Personality Disorder (APD) subject was less likely to make *victim's welfare* justifications of moral items than non-antisocial, criminal control groups. These results were in line with the position that APD represents a failure in the Violence Inhibiting Device (VIM). However, there are alternative accounts of the cause of APD that are available. This chapter will explore one such possibility; that APD is caused by a deficit in the ability to role-play.

Role playing is defined as the "imaginative transposing of oneself into the thinking and acting of another" (Feshbach, 1978). Role playing occurs when the individual assumes the attitude of another individual; when the individual forms a representation of what the other person might be thinking or feeling at that particular moment. Gough (1948) proposes that a "deficiency in role-playing means the incapacity to ... identify with another's point of view" (p. 363).

Gough (1948), proposes that the APD individual suffers from a deficit in the ability to role-play; that he is unable to identify with another's point of view. His account has been influential in the APD literature in the sense that it has been frequently cited (e.g., Gibbs, 1987; Hare, 1970; Schalling, 1978), though, admittedly, it has stimulated little research (see below). This influence is in spite of the fact that the account is little more than incoherent folk psychology; it neglects to mention mechanisms that must be assumed if the theory is to work. For example, Gough claims that because of the deficit in role-playing "the psychopath is unable to foresee the consequences of his own acts, especially their social implications, because he does not know how to judge his own behaviour from another's standpoint" (p. 364). However, this inability to judge his own behaviour from another's standpoint is only important if these judgements are necessary for the control of violent behaviour. There must be a mechanism that operates on these judgements and prevents antisocial behaviour occurring. Gough does not mention such a mechanism.

Gough proposes that the psychopath's symptomatology is a consequence of this failure in role-playing. He sketches the influence of this failure for three of the main symptoms:

> "The psychopath will violate other's wishes and desires because he does not conceive his own actions as inimical to their wants. He forms no deep attachments because he does not know how to identify himself with another or to share another's viewpoint. He lacks control because he cannot anticipate objections which others will make to his behaviour" (p. 363).

In chapter three, I developed a Developmental Contingency Model (DCM) for the Theory of Mind Mechanism (ToMM); see figure 3.2. I suggested that ToMM was a prerequisite for the development of a "Theory of Mind" and that a "Theory of Mind" was a prerequisite for role-playing. I stated that ToMM was the mechanism which allows the representation of mental states while the "Theory of Mind" was the structure which provides information on what another is feeling at a particular time. Both ToMM and a "Theory of Mind" must be prerequisites for role-playing. Role-playing involves the calculation of what another is feeling at a particular time. To role-play, the individual must have ToMM to represent the product of the activity. In addition, to role-play the individual must have a "Theory of Mind" to provide the information on what an individual thinks at a particular time. Role-playing involves the specification of the trigger conditions for the activation of a "Theory of Mind". To role-play, an individual must be setting the input conditions for the triggers of the "Theory of Mind" schema. Role-playing involves planning. Some form of plan must be set up detailing the target for the operation of ToM (i.e., the person whose mental representations are going to be calculated) and what situational features/ or memory representations (i.e., the particular details of the situation), associated with the target, will be input to the process that will eventually result in the mental state representation. However, given the purposes of this thesis, it is not necessary to develop a full description of the role-playing process.

Gough is thus suggesting that representations, formed through role-playing, result in the inhibition of violent action. I have also described a system where representations of the victim's plight inhibit behaviour. However, there is a great difference between the two approaches. This difference is in the degree of automaticity of the processes that these two approaches propose. My approach suggests that representations of the victim's distress cues automatically activate inhibitory behaviour. In terms of the Supervisory Attentional System framework (SASF) that this thesis has been embedded in, stimuli (victim's distress cues), external to the individual, having been processed by the sensory systems, pass to the trigger data base and activate an

inhibitory schema. On the other hand, role-playing within the SASF is not automatic. As I mentioned above, considerable planning is involved so that attention can be focused on stimuli relevant to the goal of the role-taking. Within the SASF framework, the S.A.S., itself, must be involved.

Of course, data already presented in this thesis raises some problems for Gough's account. The findings of chapter three reveal that the possession of a "Theory of Mind" is not a prerequisite for the development, in the individual, of the moral/ conventional distinction; autistics, without a "Theory of Mind", make the distinction between moral and conventional rules. The findings of chapter four reveal that APD subjects do not make this distinction. This finding necessarily implies that APD must involve more than a deficient "Theory of Mind". However, it might be that the APD does not even show a deficient "Theory of Mind"; there may be no role-taking deficiency in APD subjects. Indeed, following the theory presented in chapter two, this would be predicted.

However, there is some evidence (Reed and Cuadra, 1957; Widom, 1976) that has been presented in support of the role-playing position. Reed and Cuadra (1957) asked student nurses to describe themselves, others in the group, and how those others would describe them (*see* Hare, 1970). Subjects who scored high on a scale of psychopathy were those who were less accurate in their estimates of what others thought of them. They were also rated as people who were generally less insightful. In addition, Widom (1976) obtained evidence consistent with Gough's position in the repertory grids of psychopaths. She found that they did not differ from controls in their use of emotional constructs, but tended to construe more situations as "dull". Further, they failed to distinguish their own constructs as any different from those of other people.

This chapter describes a further attempt to investigate the role-playing ability of APD subjects. The test of role-taking ability used was constructed by Happé (1991). She named it the "Strange Stories" test. The test stimuli consist of a collection of stories about everyday social situations. One of these stories is presented below:

> Jill wanted to buy a kitten, so she went to see Mrs. Smith, who had lots of kittens she did not want. Now Mrs. Smith loved the kittens, and she wouldn't do anything to harm them, though she couldn't keep them all herself. When Jill visited she wasn't sure she wanted one of Mrs Smith's kittens since they were all males and she wanted a female. But Mrs. Smith said, "If no-one buys the kittens I'll just have to drown them!"

The subject is asked why the story character said what he did. When answering this question, the subject must be role-playing; putting himself into the position of the person in the story. This test was originally designed to investigate the "Theory of Mind" ability of a group of autistic subjects (Happé, 1991). This test has been validated; Happé demonstrated that performance on this task was directly correlated with performance on other "Theory of Mind" tasks.

If Gough is right and the APD subjects are deficient in their role-taking, it would be expected that they would be significantly less likely to understand why the person said what they did than the non-antisocial criminal controls. If, however, role-playing has no role in the causation of APD then there should be no relationship between group and test performance.

5.2: Method

5.2.1: Design

The experiment involved a two way repeated measures factorial design. The independent variable was the two different subject groups (antisocial personality and non-antisocial criminal controls) and the two different types of story (moral and conventional). The dependent variable was the score on a "Theory of Mind" battery.

5.2.2: Subjects

The subjects were the 10 APD and 10 Criminal Control (CC) subjects that were used in the experiment described in chapter four. See chapter four for full details of these subjects.

5.2.3: Materials

The materials consisted of a test of the subjects' "Theory of Mind". This test was the Strange Stories task developed by Happé (1991). This test involves the subject reading various stories describing naturalistic social situations and interpreting the behaviours of the story characters or what they have said or predicting how they would act. An example story is presented below:

> Katie and Emma are playing in the house. Emma picks up a banana from the fruit bowl and holds it up to her ear. She says to Katie, "Look! This banana is a telephone!"

Subjects responses to questions were recorded on standard scoring sheets.

5.2.4: Procedure

Subjects were tested in one of the interview rooms attached to whichever ward the subject was housed on. Before the study commenced the subjects were introduced to the experimenter and informed about what they were about to do. Subject consent forms were taken.

Each of the "Strange Stories" was read out to the subject one at a time. The order of presentation of the "Strange Stores" was randomized across subjects. After the "Strange Story" was over, the subject was then asked why the story character had acted as he/ she did, or why they had said what they did.

5.2.5: Scoring Procedure

The scoring procedure followed that described by Happé (1991). The subject was scored on the correctness of their response to the "Why" question. Incorrect answers might either involve errors about the facts given in the story or because it involved an inference which was inappropriate as a reason for the story character's utterance. Taking the example supplied by Happé for the story in which Anna breaks a vase but tells her mother that the dog did it, the justification "Anna did not break the vase", would be scored as incorrect because it includes a factual error. Judging appropriateness might be considered quite subjective but the very high level of inter-rater agreement obtained here (see below), and by Happé, indicated that in this context it was not.

The justifications were also scored according to whether they involved mental or physical states. For example, in the story where Emma pretends the banana is a telephone, this can be correctly explained by the physical justification "Because the banana looks like a telephone", or the mental state justification "He's just pretending". Following Happé, mental state answers included all those that referred to thoughts, feelings, desires, traits and dispositions. Mental state justifications included terms such as like, want, happy, cross, afraid, know, think, joke, pretend, lie, to fool someone, expecting. Justifications were scored as physical state when they referred to non-mental events - physical appearance, action of objects, physical events and outcomes. Physical state answers included terms such as big, looks like, is shaped like, to get rid of them,

to sell them, because of the X (object), to not get X (physical outcome; e.g., put in jail, have a filling). For these subjects, as can be seen from the inter-rater agreement scores, this distinction between mental and physical state answers was clear.

5.2.6: Co-validation of the scoring procedure

Given the possible subjectivity of the scoring procedure, co-validation of the scoring was necessary to establish validity. The justification given by every subject tested, to one story of each type, were given to a second rater. This rater was not a psychologist and was naive to the hypothesis being tested. In addition, the identity and diagnostic group of the subject was not available to the second rater. He was asked to mark each response with a tick or a cross according to whether the response provided an appropriate answer to the test question ("Why did she/ he say that?"). In addition, the rater marked the answers as either concerning mental states or physical states. She was given some examples of each type of justification, taken from the stories not included for co-validating, to illustrate these findings.

Ratings made by the experimenter and by the second rater were compared. The degree of agreement is shown in table 5.1 - which gives the percentage of agreement of justifications which were judged by both raters as correct versus incorrect, as mental versus physical, and as correct/ incorrect mental versus correct/ incorrect physical (i.e., identical ratings).

Table 5.1: The percentage of concordant ratings of justifications

Correct vs Incorrect	Mental vs Physical
97 (1.02)	95.5 (1.69)

As can be seen by the rates of agreements there were very few disagreements between the raters. In fact, there were no disagreements as regards the correctness of the justification. The few disagreements as regards the mental vs physical state of the justification were due to rater disagreements over whether such words as "rather" and "feel" should be considered mental state words. Following the guidelines of Happé, they were treated as mental state words in the following analysis.

5.3: Results

The justifications given by the subjects in response to the "Why?" question were scored as correct or incorrect, and as concerning mental or physical states. The results for the antisocial personalities and the criminal controls can be seen in table 5.2. This table clearly shows: First, that there is no real difference in the ability levels of the two groups for this task and; Second, that the error rates for both subject groups were extremely low (APD = 6.7%; CC = 8.4). A two way ANOVA performed on the data confirmed this impression; there was no significant difference in task performance between the two groups ($F(1,18)=.393$; p= n.s.).

Table 5.2: Subject's scores on the "Strange Stories" task and details of their justifications.

Group	Total Score (Max=24)	Correct Justifications		Incorrect Justifications	
		Mental	Physical	Mental	Physical
APD	22.4 (1.48)	18.7 (1.5)	3.7 (2.8)	1.3 (1.9)	0.3 (0.5)
CC	22.00 (1.22)	16.6 (4.8)	5.4 (3.8)	0.8 (1.0)	1.2 (1.6)

This lack of a difference between APD and CC subjects in their performance on the "Strange Stories" task indicates that poor role-taking cannot be the cause of APD. Comparison with Happé's (1991) group scores shows that the APD subjects scored considerably better than able autistic subjects (mean = 15.7) and only slightly poorer than her normal adult population[1] (mean = 23.7). This again indicates that APD subjects are not deficient in their ability to role-take. Indeed, in this study APD subjects appear to be marginally superior in ability.

Table 5.2 also shows the number of correct and incorrect justifications broken down by group and whether mental or physical state justifications were used. Again it is clear that APD subjects show no signs of disability relative to CC subjects; the frequency of their use of mental state justifications was not lower than that of the CCs. A two way ANOVA conducted on the data confirmed this impression ($F(1,16)=0.987$; p=n.s.). Indeed, again the performance of APD subjects was found to be marginally superior. Clearly, given these findings APD subjects cannot

[1] Happé's control group were students and thus presumably of higher IQ than the clinical population that was described in chapter four.

be thought to be impoverished in their role-taking relative to CC subjects.

5.4: Discussion

The present study examined whether the Antisocial Personality Disorder was deficient in his ability to role-take by using a "Theory of Mind" task, the "Strange Stories" task, developed by Happé (1991). This study revealed that the APD subject was not deficient in his "Theory of Mind" ability relative to CC subjects; the APD subjects judgements were correct marginally more frequently than CC subjects and they used more mental state justifications than CC subjects.

This finding complements the earlier finding (see chapter three) that autistic subjects make the moral/ conventional distinction. Chapter three and five, together, provide evidence of a double dissociation between the ability to make a moral/ conventional distinction and the ability to do "Theory of Mind" tasks. Autistics suffer from a deficient "Theory of Mind" (e.g., Baron-Cohen *et al.*, 1985) and fail false belief tasks, for example, but they do make the moral/ conventional distinction (chapter three). APD subjects do not suffer from a deficient "Theory of Mind" (this chapter) and don't make a distinction between moral and conventional rules (chapter four).

Gough's (1948) account of the origin of APD can now be dismissed; APD subjects can obviously role-take. Nor can Gough's position be saved by revising the concept of role-taking. In chapter three, I suggested that Turiel's claim that role-taking was a developmental prerequisite for the moral/ conventional distinction might be saved by such a revision. Role-taking (forming a representation of the target individual's mental representation) cannot be a developmental prerequisite for the moral/ conventional distinction since autistics who clearly lack the ability to perform this type of role-taking do make the distinction (Chapter three). However, role-taking (attempting to access similar past experiences that have occurred to the self) might be a developmental prerequisite for the formation of the moral/ conventional distinction. Autistics might be able to accomplish this form of role-taking and thus Turiel's claim is not disproved. Such an argument in favour of Gough cannot work here; there is no reason to believe that the APD is unable to access similar past experiences that have occurred to the self.

The data presented in this chapter contradicts the findings of earlier studies (Reed & Cuadra, 1957; Widom, 1976). However, this should not be thought problematic; the Reed and Cuadra

(1957) and Widom (1976) findings are suspect. Both studies identified their "psychopath" group through methods of doubtful validity (see Hare, 1985). In addition, the Reed and Cuadra (1957) study was conducted on student nurses. The subjects were at either end of the a measure of psychopathy. However, it seems unlikely that either group were "psychopaths". The psychopathy scale itself may actually only have measured "Theory of Mind" ability. The "psychopaths" may just have had poorer "Theory of Mind" competence and this may have been picked up by both the psychopathy scale and the task.

5.5: Conclusions

In this fifth chapter, I have described an alternative account of APD; that it is caused by an inability to role-take. The following conclusions can now be drawn:

(1) The theory, as it was described by Gough (1948), is massively under-specified. Additional mechanisms need to be postulated to make the account work.

(2) The theory is wrong. APD is not associated with impoverished role-taking ability relative to controls.

In summary, it can be concluded that any account of APD which suggests that impoverished role-taking is the cause of the disorder can be dismissed.

Chapter Six

Antisocial Personality Disorder: A Lack of Inhibition?

6.1: Introduction

In chapter four, I suggested that the symptoms which make up Antisocial Personality Disorder (APD) were developmental consequences of the absence of a Violence Inhibition Mechanism (VIM). In chapter five, I investigated an alternative account of APD proposed by Gough (1948). According to Gough, APD is a consequence of an inability to role-take. However, the evidence presented in chapter five demonstrates that this account is incorrect; relative to controls, APD subjects are not impaired in role-taking.

In this chapter, I will investigate a second alternative account of APD; that APD results from damage to the Supervisory Attentional System (SAS). Various authors have advocated this position, notably Newman, his colleagues and Gorenstein (Gorenstein, 1982; Newman & Kosson, 1986; Newman & Gorenstein, 1980; Newman, Patterson & Kosson, 1987; Newman, Widom & Nathan, 1985) though none has made explicit reference to the S.A.S. Instead, there have been references to the inability of APD patients to shift "response sets" (Newman *et al.*, 1985; 1986; 1987) and references to the similarity in symptomatology of APD subjects and patients who have suffered damage to their frontal lobes (Gorenstein, 1982). Both of these positions, and the empirical evidence associated with them, will be described below (sections 6.1.1 and 6.1.2). Following this, both of these positions will be incorporated within the S.A.S. framework.

6.1.1: The "response set" position of Newman *et al.*

During the sixties and seventies, several theorists considered that moral socialisation was achieved through punishment (e.g., Eysenck, 1964; Trasler, 1978). These theorists claimed, for example, that remorse was a consequence of conditioning; "Conscience is a conditioned reflex" (Eysenck, 1964; p. 64). Their basic position was that anxiety, induced by the transgression being punished, becomes associated with the transgression. This argument was extended to the claim that APD was a consequence of a deficiency in conditionability. While the central thrust

of this argument has been discredited,[1] it did generate a literature investigating the learning abilities of the APD subject. For example, Lykken (1957) observed that APD subjects were significantly inferior to normal controls and a "neurotic sociopathic" group in attaining a conditioned galvanic skin response to a buzzer that had previously been paired with an electric shock. In addition, Lykken (1957) observed differences between APD subjects and controls on a maze learning task. This task involved the subject having to choose the lever out of four alternatives which would advance him through the maze while learning to avoid another lever out of the four alternatives which if pressed would result in shock. Lykken observed that the APD subjects did not differ from controls in their ability to learn their way through the maze but that they were significantly poorer at learning to avoid punishment. However, Schmauk (1970), replicating this study, found that APD subjects were only significantly poorer at learning to avoid electric shock punishment; if the punishment was tangible (i.e. loss of money) they learnt as fast as controls. Schmauk interpreted his finding as indicating that monetary punishment was more effective than electric shock in mediating avoidance learning in APD subjects because of the "different values of the psychopath". Schmauk suggested that APD subjects were not aversively aroused by electric shock punishment but that they were by monetary punishment.

The "response set" position of Newman and his colleagues can be considered a reinterpretation of Schmauk's data. Newman and his colleagues explain Schmauk's data in terms of the "saliency" of the task requirements. They suggest that making the punishment tangible makes it more salient and prevents/ ameliorates "response set perseveration". They refer to McCleary (1966) who has defined perseveration as the tendency to continue a response set for reward despite punishment or changes in environmental contingencies that reduce the adaptiveness of continued responding. However, the data they are attempting to explain (and to generate: Newman *et al.*, 1985; Newman & Kosson, 1986) cannot be interpreted in terms of this definition of perseveration. They are referring to what they term "response set perseveration"; the inability to learn to gain reward and avoid punishment simultaneously (see section 6.1.3 for a contrast between perseveration and "response set perseveration" within the SASF).

Newman *et al.*, (1985) tested their position by giving APD delinquent subjects and non-APD

[1]Developmental studies have clearly demonstrated that punishment is not associated with a reduction in antisocial behaviour (*see*, for a review, Hoffman, 1977). Instead, a reduction in antisocial behaviour has been associated with inductive techniques (e.g., asking the child how the victim of the act feels).

Table 6.1: The task conditions for the Newman *et al.*, (1985) study.

	Stimuli	Desired response	Consequences if respond	Consequences if do not respond
Paradigm 1	1-4 (S+s)	Respond	Reward	--
	5-8 (S-s)	Do not respond	Punishment	--
Paradigm 2	1-4 (S+s)	Respond	Reward	--
	5-8 (S-s)	Do not respond	--	Reward

delinquent subjects two paradigms (see table 6.1 for paradigm details). Both paradigms involved the presentation of eight two-digit numbers ten times in different randomized orders for a total of eighty trials. The subjects had to learn to make responses to four of these stimuli (the S+'s) and not to make responses to the other four of these stimuli (the S-'s). In paradigm one, the subject was rewarded for responding to S+'s and punished for responding to S-'s. In paradigm two, the subject was rewarded for responding to S +'s and rewarded for not responding to S-'s (see table 1 for experiment details). The APD delinquent subjects responded inappropriately to the S-'s significantly more often than the non-APD controls when they were given the competing goals of avoiding punishment as well as earning reward than when they were only earning reward.

Table 6.2: The task conditions for the Newman and Kosson (1986) study.

	Stimuli	Desired response	Consequences if respond	Consequences if do not respond
Paradigm 1	1-4 (S+s)	Respond	Reward	--
	5-8 (S-s)	Do not respond	Punishment	--
Paradigm 2	1-4 (S+s)	Respond	--	Punishment
	5-8 (S-s)	Do not respond	Punishment	--

This result has been replicated and extended by Newman & Kosson (1986). In their study, adult APD subjects and non-APD controls did one of two paradigms (see table 6.2 for paradigm details). In paradigm 1, the subjects received rewards for responding to S+'s and punishments for responding to S-'s. In paradigm 2, the subjects received punishments for not responding to S+'s and punishments for responding to S-'s. The adult APD subjects responded inappropriately

to the S-'s significantly more often relative to controls when they were given the competing goals of avoiding punishment and earning reward than when they were only being punished.

While Newman *et al.*, (1985) and Newman and Kosson (1986) demonstrated "response set perseveration" in APD subjects relative to controls, Newman *et al.*, (1987) showed increased perseveration in APD subjects. Newman *et al.*, (1987) gave APD subjects and less antisocial controls a computerized card playing task. Subjects were each given 10 chips worth 5 cents. Subjects were asked before each trial whether they wanted to play (the question DO YOU WANT TO PLAY? appeared on the screen). Playing involved pressing a button and this would result in the appearance of a card on the screen. If this card was a face card (J, Q, K, A) the subject was rewarded with an extra chip. If this card was a number card, the subject lost a chip. With each block of ten trials the probability that the subject would receive a number card (i.e., be punished) increased. Subjects were placed in one of three conditions. In condition I, subjects could respond to the next card as quickly as they wished and received *immediate* feedback only (i.e., a computer-delivered message and the addition, or subtraction, of a chip). In condition IC, everything was the same except that the subjects also received *cumulative* feedback; after each play, the letter or number that had appeared was rewritten at the top of the monitor. These letters and numbers appeared in rows of 10 across and remained visible throughout the task. In condition ICW everything was the same as in condition IC except that the question DO YOU WANT TO PLAY? did not appear for 5 seconds after the previous play. APD subjects played significantly longer, they perseverated (and lost more money), than did less antisocial controls in conditions I and IC. However, this difference was not present in condition ICW. Newman *et al.*, (1987) argue that in this last condition the 5s delay prevents perseveration.

6.1.2: The frontal lobe position of Gorenstein

There is good evidence that APD is not associated with any form of global brain damage. Hart, Forth and Hare (1990), using a variety of neuropsychological tests, demonstrated that the APD was not associated with any generalized impairment. However, links have been repeatedly made between APD and damage to the frontal lobes (Elliot, 1978; Gorenstein, 1982; Newman & Gorenstein, 1980; Schalling, 1978).

Four clusters of behavioral problems have been associated with frontal damage (Lezak, 1976). These are:

1) Slowing - apathy, loss of initiative or ambition;

2) Perseveration - failure to suppress on-going activities when they are no longer appropriate;

3) Deficient self-awareness - the inability to perceive performance errors, inability to appreciate the impact one makes on others, tendency toward self-satisfaction, little or no anxiety, impulsiveness, lack of concern about social conventions;

4) Concrete attitude - stimulus bondedness, inability to plan or sustain goal directedness.

Gorenstein (1982) drew attention to what he thought was the similar symptomatology of APD subjects and patients with damage to their frontal lobes. He noted that of the four clusters of behaviour problems associated with frontal damage described above, three can be found ("slowing" is not associated with APD) in the clinical description of APD.

Unfortunately, those theorists who have suggested that APD might be the consequence of a dysfunction in the operation of the frontal lobes have not fully described the psychological implications of such a dysfunction. Gorenstein (1982) has stated that perseveration associated with frontal lobe impairment may be the cause of the APD's inability to deal with competing demands (indicated by the findings of Lykken (1957), Schmauk (1970) and Newman and his colleagues). However, Gorenstein has not explained how this tendency to perseverate might give rise to the specific behaviours that make up APD.

6.1.3: The Supervisory Attentional System Framework (SASF) and APD

The SASF has already been described in this thesis (see chapter two). This chapter will describe how damage to the S.A.S. might give rise to antisocial personality disorder. A possible explanation has already been provided by Shallice (1986). Shallice (1986) claimed that one of the consequences of damage to the S.A.S. would be an inability to resist temptation; i.e., an inability to inhibit a prohibited response [schema] when a stimulus eliciting that response is present in the environment. Shallice considers that highly tempting stimuli are stimuli which, when passed onto the trigger data base, immediately activate a behavioral schema. To resist temptation it is necessary for the S.A.S. to inhibit the activated schema. Damage to the S.A.S.

would prevent this inhibition and, consequently, the schema activated by the stimulus would be activated; there would be no resistance to temptation.

Shallice (1986) also predicts increased perseveration as a consequence of SAS damage. Within the SASF, perseveration can be conceptualised as a failure of the SAS to deactivate the schema that had been determining behaviour. Thus, the observation of increased perseveration in APD subjects by Newman *et al.*, (1987) would be explainable in terms of SAS damage. However, the Newman *et al.*, (1985) and Newman and Kosson (1986) findings of "response set perseveration" in APD subjects are not so easily accountable. "Response set perseveration" implies the inability to learn to avoid punishment and to gain reward simultaneously. Within the SASF, "response set perseveration" can be conceptualised as an inability to set up two schemata simultaneously (one schema responsible for gaining reward and the other for inhibiting punishment). Of course, this conceptualisation assumes that schemata are set up according to response conditions and that the SAS is involved in the formation of schemata.

In contrast to the Newman's ideas, Gorenstein's noting of the similarity in symptomatology of APD and frontal lobe patients necessitates that there must be widespread damage to the S.A.S. It would not be possible to propose that the damage to the S.A.S. has been confined to the S.A.S. mechanism which interacts with the trigger data base (i.e., the S.A.S. mechanism that is responsible for schema activation and prevents perseveration). While "perseveration" and "deficient self awareness" (as Gorenstein conceptualises them) might be consequences of damage to the sub-component of the SAS which mediates contingency scheduling, the third stimulus cluster ("concrete attitude") which Gorenstein also associates with APD cannot be. "Concrete attitude" involves the inability to plan. Shallice and Burgess (1991; *in press*) have suggested that there is a sub-component of the SAS responsible for planning. Thus, a psychological interpretation of Gorenstein's position must assume damage to at least two of the components of the SAS; schema mediation and planning.

6.1.4: Tests of the executive functioning of APD subjects

Various investigations of the executive functioning (i.e., SAS functioning) of criminal and APD subjects have been attempted. For example, Pontius and Yudowitz (1980) found that 36 percent of their criminal subjects doing the Narratives Test and 33 percent of their criminal subjects on Trail Making Test-B performed at a level indicative of frontal lobe dysfunctioning. There was

a very significant association between the subjects' performances on these tasks. This study is only of limited interest, however, because it examines the criminal population generally and not APD in particular. Also there was no control group.

Three studies have applied batteries of executive function tests to delinquent subjects. Skoff and Libon (1987) compared the scores of 22 incarcerated delinquents to published test norms for the WCST, Porteus Mazes, Trails B, Verbal Fluency, and four additional executive tasks. They reported that one third of their subjects scored in the impaired range on the battery as a whole. This study suffers from the absence of a non-delinquent comparison group; the test norms are of questionable relevance for the delinquents of California, who might be expected to be low SES minority group members. Appellof and Augustine (1985; cited by Moffitt and Henry, 1989) tested 30 male delinquents and 30 controls (selection criteria not described) on the WCST, Porteus Mazes, Verbal Fluency, and six other unnamed measures. They found no group differences and concluded that executive dysfunctions do not differentially characterise delinquents. Moffitt and Henry (1989) divided subjects into four groups on the level of their self reported delinquency level and comorbidity with Attention Deficit Disorder (ADD). These subjects were tested on WISC-R mazes, the WCST, Trail Making Test A and B and the Rey-Osterreith Complex Figure Test (ROC). The delinquent subjects who also gave evidence of ADD were found to have significantly more impaired performance on these executive function tasks than the other three groups. Again though, these studies concentrated on the delinquent population generally and not APD subjects in particular.

Three studies have looked at the executive functioning of the adult antisocial personality. Gorenstein (1982) compared 20 male psychopaths to 23 control subjects taken from the same hospitals and to 18 male college students. Tests used were the Wisconsin Card Sorting Test, the Stroop Colour-Word Interference Test, the Sequential matching memory test and the Necker cube test. APD subjects made more perseverative errors on the Wisconsin Card Sorting Task, more errors on the sequential memory matching task and more Necker cube reversals than the control group. Hare (1984) has questioned, however, if the psychopathic group was really made up of psychopaths. The division into the two groups was made according to the results obtained from two self-report measures: a self-report version of the Research Diagnostic Criteria (Spitzer, Endicott, & Robbins, 1975), for which no information on reliability or validity was offered; and the Socialization scale (Gough, 1969), which was only used if the subject did not score at least twice the number of symptoms needed to satisfy three of the four criteria used to diagnose

psychopathy. No reliability quotients were reported for the independent validation used - ratings by psychiatric residents.

Hare (1984) attempted to replicate Gorenstein's findings. The tasks used were the Wisconsin Card Sorting task, the Necker Cube and a sequential memory test. Though there was a trend for the APD subjects to make more perseverative errors on the Wisconsin Card Sorting task this did not prove to be significant. In addition, Hoffman, Hall, & Bartsch (1987) attempted a replication of the Gorenstein findings and they again found no relationship between psychopathy and executive functioning.

There is thus some confusion in the literature over the executive functioning of APD subjects. While it seems clear that there is an association between poor executive functioning and criminal/ violent behaviour it is much more debatable whether there is an association between poor executive functioning and APD. This lack of clear evidence is especially apparent if the tests that have been used with APD subjects are considered. The Wisconsin Card Sorting Task has been extensively criticised (Burgess, 1992), the Necker cube task seems intrinsically unreliable and no validity data is presented for the Sequential Memory Test. While the validity of the Stroop Colour-Word Interference Test is reasonably established, this task was only used in one of the three studies. Because of this lack of clear evidence and because of the criticisms of the tasks that have been previously used it was decided to again investigate the executive functioning of APD subjects. Four tasks were selected: two ostensibly of perseveration/ response suppression (Verbal fluency and the Hayling task) and; two ostensibly of planning/ set formation (cognitive estimates and the Brixton task).

6.2: Method

6.2.1: Design

The experiment involved a two way repeated measures factorial design. The independent variable was the two different subject groups (antisocial personality and non-antisocial criminal controls). The dependent variable was the score achieved by the subjects on the various executive function tests.

6.2.2: Subjects

The subjects were the same as those used in the study described in chapter four (see table 4.1 for subject characteristics) except that one of the APD subjects was not administered the cognitive estimates or the Brixton tests.

6.2.3: Materials

The tests of executive functioning used were: the Verbal Fluency test (Milner, 1964); the Hayling task (Burgess, 1992); the Cognitive Estimates test (Shallice & Evans, 1978) and; the Brixton task (Burgess, 1992).

Verbal Fluency: This task involves the experimenter giving the subject a letter. The subject must then say as many words as he can that begin with that letter in a 60 second period. The letters used were F, M and S. Order of presentation of letters was randomized across subjects.

The Hayling task: This task occurs in two stages. In the first stage (A), the subject is asked to complete sentences, that are read out by the experimenter, with a single word. The word used by the subject and the time taken are recorded. In the second stage (B), the subject is asked to complete the sentence read out by the experimenter with a word that does not make sense in the context of that sentence. The subject is scored according to how related his response is to the sentence. For example, one sentence read to the subject is "The Captain stayed with the sinking ----". If the subject makes a straight completion (i.e., ship) to this sentence he scores 3, if he gives a semantically related response (e.g., plane) he scores 1, and if he makes a completely unrelated response (e.g., wall) then he scores zero.

Cognitive Estimates: This task involves the subject being asked ten questions (e.g., What is the height of the Post Office tower?) which he is unlikely to be able to answer from general knowledge. Instead, the subject must develop some form of estimation strategy to work out the answer.

The Brixton task: This task involves the subject being presented with a book of stimuli. Each stimulus page has ten circles drawn on it. One of these circles is coloured in. The subject must work out which circle will be coloured in on the following pages. The identity of the coloured-

in circle is rule bound. To achieve correct performance, the subject must work out what the rule is and use this to predict which circle will be coloured in on the following pages. The rule determining the identity of the coloured in circle changes over time. The subject must identify when the rule has changed and determine what the new rule is.

Subjects responses to questions were recorded on standard scoring sheets.

6.2.4: Procedure

Subjects were tested in one of the interview rooms attached to whichever ward the subject was housed on. Before the study commenced the subjects were introduced to the experimenter and informed about what they were about to do. Subject consent forms were taken. When they were comfortable experimenting began.

The order of the executive function tasks was randomized across subjects.

6.3: Results

The subjects were scored according to their performance on each of the four executive functioning tasks used in this study. The scores of the two subject groups on these four tasks are displayed in table 6.3. As can be seen in this table the performance of the two groups was remarkably similar. There was no evidence that antisocial personality subjects had more difficulty with these tasks than criminal controls. Indeed, as can be seen from the figures there was a slight tendency for the antisocial subjects to perform at a higher level than the criminal controls.

Two way ANOVAs (comparing the two subject groups) were conducted for each of the executive tasks submitted to the subject groups. These demonstrated that there was no significant difference in the performance of these subject groups on: the Verbal Fluency task ($F(1,18) = 1.043$; n.s.); the Cognitive Estimates task ($F(1,17) = 0.180$; n.s.); and; the Brixton task ($F(1,17) = 0.612$; n.s.). Two way ANOVAs (comparing the two subject groups) were also conducted on the two measures generated by the Hayling task. The APD subjects did not perform stage A significantly slower ($F(1,18) = 0.294$; n.s.) or score higher on stage (B) ($F(1,18)= 0.047$; n.s.). After these results it can be safely concluded that the antisocial

personality cannot be distinguished from a criminal control group on the basis of his executive functioning.

Table 6.3: Performance of APD and CC subjects on executive function tasks (VF = Verbal Fluency; two measures from the Hayling task; time to read stage A and score on stage B; CE = Cognitive Estimates; the Brixton task).

	Executive function tasks				
	VF	Hayling A	Hayling B	CE	Brixton
APD	44.40 (13.57)	17.25 (7.77)	10.00 (6.00)	7.11 (5.46)	19.38 (7.23)
CC	38.00 (10.48)	19.32 (8.71)	10.56 (5.08)	8.80 (4.61)	22.60 (9.67)

Table 6.4: Performance of anterior and posterior patients and normal subjects on executive function tasks (VF = Verbal Fluency; two measures from the Hayling task; time to read stage A and score on stage B; CE = Cognitive Estimates; the Brixton task).

	Executive function tasks				
	VF[2]	Hayling A[3]	Hayling B	CE[4]	Brixton[5]
Anterior	18.9 (7.5)	28.6 (34.5)	11.5 (11.3)	8.19 (4.28)	24.5 (8.9)
Posterior	28.4 (11.0)	17.4 (14.1)	4.8 (4.8)	4.71 (3.91)	18.3 (7.2)
Normal controls	34.8 (14.3)	14.8 (12.5)	4.7 (5.5)	3.6 (1.9)	18.2 (4.3)

In table 6.4, I provide data from anterior and posterior patients and normal control groups. This data indicates that for most of these tasks neither the APD nor the CC groups were performing well. Only on verbal fluency were both groups performing better than expected. For the Hayling A score, both APDs and CCs were within one standard deviation (s.d.) of the normal

[2] The data is taken from Miller (1984). Both the anterior and posterior patients had suffered left hemisphere damage.

[3] The data for both Hayling measures is taken from Burgess (1992).

[4] The data is taken from Shallice and Evans (*personal communication*). Both the anterior and posterior patients had suffered left hemispheric damage.

[5] The data is taken from Burgess (1992).

controls in the Burgess (1992) study. However, the Hayling B scores for both groups were high; the CCs scoring more than one s.d. above the mean of the normal controls in the Burgess (1992) study. On the cognitive estimates task, both the APDs and the CCs were performing more than one s.d. above the mean of the normal controls in the Shallice and Evans study. Indeed, the CCs were performing worse than the anteriors in the Shallice and Evans study. However, following the methodology adopted by Delbecq-Derouesné *et al.*, (1990), it was found that the CC scores were performing within the normal range for their age and IQ.[6] On the Brixton task, the CCs, but not the APDs, were performing more than one s.d. above the mean of the normal controls in the Burgess (1992) study. However, again the low IQ of the subjects in this experiment should be considered; the average IQ of the subjects in the Burgess study was 103. Burgess (1992) has found IQ to significantly predict performance on the Brixton task.

In table 6.5, I show the percentage of subjects in each group whose performance in the two perseveration/ response suppression tasks (Verbal fluency and the Hayling B score) and the two planning/ set formation tasks (Cognitive Estimates and the Brixton task) who are not impaired according to the cut-off scores for these tasks. The cut-off scores for Verbal Fluency and the Brixton task were age and IQ weighted. However, since the B score of the Hayling task has been found to be relatively age and IQ independent (Burgess, 1992), the advised cut-off score of 10 was adopted (Burgess, *personal communication*). Since the cut-off score of the Cognitive Estimates task has not yet been age and IQ weighted the accepted cut-off level (8) was adopted here.

Table 6.5: The percentage of subjects in both groups found to be performing better than the cut-off points (i.e., normally) for each of the tasks (VF = Verbal fluency; CE = Cognitive Estimates).

	Executive function tasks			
	VF	Hayling B	CE	Brixton
APD	90	40	90	90
CC	80	50	80	80

[6] A regression line was derived from the data on 84 normal controls taken from McKenna, Burgess and Shallice (*in preparation*). This predicted a cognitive estimates score of 6.8 for a population with the age and IQ of the CCs. The actual score of the CC subjects, 8.8, fell within the standard error (3.66) of the regression equation. Thus, it can be concluded that the CC subjects were performing within the normal range for their age and IQ.

As can be seen, a similar proportion of APD and CC subjects performed within normal limits in these tasks. Indeed, a chi-squared test on the number of subjects performing better than the cut-off levels revealed no significant difference between the groups ($X^2 = 0.45$, d.f. = 3, p = n.s.). As can also be seen, the APDs were generally performing within normal limits on all of the tasks except the Hayling; 6 of the APDs Hayling B scores exceeded the cut-off level. Most of the CC subjects were performing within normal limits on Verbal Fluency and the Brixton task. However, half of these subjects were performing at a level lower than the cut-off point on the Hayling task and Cognitive Estimates.

In summary, it is clear that the APD subjects were not significantly poorer in their executive functioning than the CC subjects. It also seems that most of the APD subjects, on all of the tasks except the Hayling B, were performing within normal limits.

6.4: Discussion

The present study investigated whether the executive functioning (the functioning of the S.A.S.) of the APD subject was deficient relative to criminal controls. Four tasks were used (the Cognitive Estimates test (Shallice & Evans, 1978); the Verbal Fluency test (Miller, 1984); the Brixton task (Burgess, 1992) and; the Hayling task (Burgess, 1992). This study revealed that the APD subjects were not deficient in executive functioning relative to CC subjects; the APD subjects actually performed marginally better than the CC subjects on all the tests performed. Comparison with data obtained previously from patients and normal controls revealed that most of the APD subjects were performing within normal limits on all of the tasks except the Hayling.

These findings have implications for both of the accounts of APD described above; Newman and his colleagues' "response set" position and the "frontal lobe" position of Gorenstein (1982) and others. I will describe the implications for the "frontal lobe" position, first.

The tasks used in this experiment were "frontal lobe" tasks (tasks that patients with damage to the frontal lobes find difficulty on). Using these tasks, no evidence of frontal lobe impairment relative to the CCs was observed. Coupling this observation with the studies of Hare (1984) and Hoffman *et al.*, (1987), it can be safely concluded that the APD of these patients was not caused by damage to the frontal lobes. Gorenstein's (1982) position, as far as it can be measured through these executive function tasks, can be dismissed.

In section 6.3, I developed Gorenstein's account within the S.A.S.F. These findings indicate that the APD patient does not have an S.A.S. impairment relative to CC controls. In addition, the APD subjects developed cognitive strategies as well as did controls; 80% of the subjects were within normal limits on the Cognitive Estimates and the Brixton tasks. It can safely be assumed that cognitive strategies must be generated for competent performance on these tasks. As regards their tendency to perseveration, one of the tasks showed that there was no evidence of increased perseveration, relative to normal controls, in APD subjects (Verbal Fluency), the other that there was (the Hayling B score). However, it should be noted that the tendency to perseverate was no greater than that observed in the CC subjects; the APD subject does not have an S.A.S. impairment relative to CC subjects.

Newman and his colleagues (Newman *et al.*, 1985; Newman & Kosson, 1986) have reported increased response set perseveration in APD subjects relative to *criminal controls*. The findings presented here go against this claim. The APD subjects did not show a greater tendency to perseverate than the CCs. However, the APDs (and the CCs) did show a greater tendency to perseverate than normal controls. There are several possible interpretations of this second finding:

(1) The increased perseveration of the APD plays no role in the development of the disorder. The mechanism which, when damaged, results in the development of the disorder (according to the present account, VIM) might be located in the brain close to the mechanism which, when damaged, results in increased perseveration. The damage resulting in the disorder might be non-specific and also damage those mechanisms responsible for the successful completion of the Hayling task.

(2) APD is a developmental consequence of the increased perseveration/ reduced response suppression revealed by the Hayling task. The damage resulting in the increased perseveration/ reduced response suppression of the CCs occured after the developmental window for APD. Of course, this account does not explain why the APDs fail to show the moral/ conventional distinction (*see* Chapter four). Nor does this account explain the APD's absence of remorse/ guilt or empathy.

(3) APD is a developmental consequence of increased perseveration/ reduced response suppression and damage to an additional cognitive mechanism. Individuals who only show increased perseveration/ reduced response suppression (i.e., the CCs) do not develop the disorder. Potentially, it is possible that individuals with damage only to the additional cognitive mechanism also do not develop the disorder. However, this latter claim is, as yet, untested.

(4) The APDs in the Special Hospitals tested in this study are a subpopulation of patients with APD. These patients may have been admitted to the Special Hospitals because of symptoms consequent on their increased perseveration/ reduced response suppression.

These findings also have implications for theories of the development of morality. Several theorists have stressed the role of "reasoning" in the development of morality. Kohlberg (e.g., Colby, Kohlberg, & Kaufman, 1987) proposed that the transition from one level to another was a consequence of reasoning. Turiel, Smetana and Nucci have all referred to self-constructional techniques as the origin of the moral and conventional domains that give rise to the moral/ conventional distinction. Turiel (1983) stated that counter-factual reasoning (generating representations of events that do not exist at the present time) might be important in the moral/ conventional distinction. According to Turiel, the child will arrive at "judgements of moral necessity" through comparisons of the performance of the act itself with its opposite. If the constructed consequences of its non-occurrence (there is no victim) are judged to be more "desirable" than the consequences of its occurrence (the victim is harmed), then inferences will be made regarding how people should act in these circumstances. According to Turiel, this same process when applied to conventional transgressions, will not result in automatic prescriptions. Turiel argues that comparison of a conventional act with its opposite will not result in one situation being obviously superior to the other. According to Turiel (1983), it is social organizational factors, such as consensus, rules, and authority that provide meaning to conventional prescriptions. Following this argument, it might have been argued, if the APD had shown impoverished SAS functioning relative to criminal controls, that it was this impoverishment that gave rise to the failure of the APD subjects to make the moral/ conventional distinction. However, the APD subjects were not found to be impoverished in their SAS functioning relative to CCs. The failure of APD subjects to make the moral/ conventional distinction must still be explained.

6.5: Conclusions

In this fifth chapter, I have described several alternative account of APD which can be interpreted within the SASF as suggesting that the cause of the disorder is damage to the SAS. I have described an experiment assessing SAS functioning in APD and criminal controls. Increased damage to the SAS in APDs relative to CCs was not found.

Chapter Seven

Antisocial Personality Disorder: A Lack of VIM?

7.1: Introduction

Chapter five and six investigated two rival positions to the suggestion that Antisocial Personality Disorder (APD) is a consequence of a deficient Violence Inhibition Mechanism (VIM). These two alternative hypotheses were: that APD was a consequence of an inability to role-take (chapter five) and; that APD was a consequence of a deficiency in the Supervisory Attentional System (chapter six). Evidence presented in these two chapters revealed that APD subjects were not impaired, relative to controls, in their ability to role-take or in the functioning of their Supervisory Attentional System. These two alternative accounts can therefore be discarded.

Chapter seven will reconsider the implications of a deficient VIM to the APD subject. More specifically, chapter seven will test a hypothesis generated from the VIM position; that APD subjects will attribute guilt differently from Criminal Controls (CCs).

In chapter four, I developed a causal model showing the developmental consequences of a lack of VIM (see figure 4.2). These were: a lack of the moral emotions (sympathy, guilt and remorse); a lack of the inhibition of violent action; a lack of the moral/ conventional distinction and; a lack of empathy. In chapter four, I pointed out that APD was defined by the lack of inhibition of violent action and I showed that APD subjects failed to distinguish moral and conventional transgressions. In this chapter, I will investigate the moral emotions of the APD subject.

As stated in chapter four, APD has long been linked to a lack of the moral emotions (particularly remorse and guilt) and empathy.[1] Karpman (1941) reported an absence of empathy in the psychopath. Items on Hare's psychopathy scale (Hare, 1980) include a lack of remorse or guilt and a tendency to be callous and to lack empathy. The DSM-III-R category, antisocial personality disorder, states that there is an absence of remorse (DSM-III-R, APA, 1987). Early theorists (e.g., Eysenck, 1964) claimed that APD subjects were less likely to feel guilt (because

[1]Please note that empathy is a moral emotions in the sense that it is a product of arousal generated by the activation of VIM. However, empathy is distinguished from the other moral emotions because certain triggers for VIM are prerequisites for its development; VIM must be activated by representations formed through role-taking.

guilt was due to conditioning and APD subjects were less "conditionable"). More recently, Gibbs (1987) has suggested a lack of empathy as the cause of the disorder. However, despite the theoretical claims and clinical descriptions there have been no empirical demonstrations of a lack of moral emotions or guilt in APD subjects. The clinical descriptions may observe that APD subjects are less likely to state, or demonstrate, feelings of guilt/ empathy about the acts they have committed. However, these observations do not necessarily indicate that APD patients lack the feeling of guilt/ empathy only that they rarely make reference to this state.

At present, there are no direct tests of whether a person feels guilt (though see chapter eight for a proposed direct test of the presence of VIM). The emotional capacities of the APD subject were therefore investigated through indirect measures; the examination of emotion attributions. Emotion attributions are here defined as the emotion that a subject believes that another is feeling.

There is a considerable literature on the development of emotion attributions in children. There is clear evidence that young children have a fairly sophisticated conception of various emotions such as happiness, sadness, anger and fear (Fabes, Eisenberg, McCormick & Wilson, 1988; Gnepp, McKee & Dominic, 1987; Strayer, 1986; Taylor & Harris, 1983). For example, happiness is associated with achievement conditions (Strayer, 1986). In addition, Bennett (Bennett, 1989; Bennett & Gillingham, 1990) has investigated the development of children's attributions of embarrassment. Bennett (1989) found that children from the age of five will attribute embarrassment to an individual who is suffering unpleasant social consequences (laughing, teasing) following an act that results in attention to the self. Children from the age of eight attribute embarrassment to any character who has committed an act which results in attention to the self even if the audience is passive.

Despite this descriptive data on the developing capacities of the child to attribute emotion states correctly, there have been no accounts of the determinants of this capacity. However, it seems plausible to suggest that the emotion attributions of the individual are, under normal circumstances, a function of the emotional experiences of the individual; i.e., if an individual is not able to experience a particular emotion, their attributions of this emotion may be anomalous. As suggested above, APD subjects lack VIM. The activity of VIM in normally developing individuals results in an aversively experienced arousal state (*see* chapter two). This arousal state can be experienced as guilt, remorse or sympathy. Without VIM, APD patients would not

be expected to experience these emotions. It might, therefore, be expected that APD subjects would differ from CCs in their attributions of guilt.

Assuming that emotional experience is reflected in the emotion attributions of the subject, analysis of the emotion attributions of the APD subject allows an alternative hypotheses to be tested. In chapter four, I presented a causal model (figure 4.3) showing the consequences of the absence of a factor "X" which was assumed to be a prerequisite for the development of all emotions. Referring back to the model of emotion presented in chapter two (section 2.4.2), it can be speculated that factor "X" might be the inter-schema competition process or the "arousal schema". A developmental consequence of the absence of either of these mechanisms would be a lack of all emotional experience. A lack of all emotional experience should presumably result in the APD subjects differing from CCs in their attributions of all emotions.

7.2: Method

7.2.1: Design

The experiment involved a two way repeated measures factorial design. The independent variable was the two different subject groups (antisocial personality and non-antisocial criminal controls). The dependent variable was the emotion attribution the subject gave for each of the stories.

7.2.2: Subjects

The subject characteristics have already been described in chapter four (see table 4.1 for subject details). All the subjects involved in the previous experiment took part in this experiment.

7.2.3: Materials

The stories used to measure the emotion attributions of the subjects were all generated from previous research. Eleven "emotion" stories were devised; stories designed to elicit attributions of particular emotions (happiness, sadness, embarrassment and guilt). Achievement conditions are associated with happiness attributions (Strayer, 1986). There was one "happiness" story (an individual wins the pools) and one "sadness" story (a person comes last in an art competition).

Embarrassment attributions are determined by audience conditions (Bennett, 1989). Following Bennett, there were three forms of audience condition: no audience; passive audience and; negative audience. Acts placed within the audience conditions were: falling out of a car; dropping a tray of food and; knocking over a painting.

Guilt has been defined as "a combination of empathic and sympathetic distress and a self-blame attribution" (Hoffman, 1987; p. 55). The six "guilt" stories were designed according to this definition. They were divided into three groups: two "intentional harm stories", subdivided into person-harm (a man punches another man) and object-harm (a man smashes up public property) and; two "unintentional harm stories", subdivided into person-harm (a man throws a stone at a bottle on a wall, it misses the bottle, flies over the wall and hits someone) and object-harm (a man driving his friends car skids on some oil and destroys the car) and two conventional stories (two children talk in class during a lesson; a child walks out of a class during a lesson).

Subjects responses to questions were recorded on standard scoring sheets.

7.2.4: Procedure

Subjects were tested individually in one of the interview rooms attached to whichever ward the subject was housed on. Before the study commenced the subjects were introduced to the experimenter and informed about what they were to do. Subject consent forms were taken.

The attribution stories was read out to the subject one at a time. The order of presentation of the stories was randomized across subjects. The acts for the audience conditions in the "embarrassment" stories were randomized across subjects. All subjects received all of the stories. After the story had been presented, the subject was asked:

> How do you think that person [referring to the story protagonist] would feel in that situation?

All responses were recorded by hand on a standard scoring sheet.

7.2.5: Scoring procedure

The subjects' responses for all the stories (except the "guilt" stories) were scored according to whether the subject gave the expected response or not; i.e., if the subject gave a happiness response to a "happiness" story this was scored, any other response to this story would be scored as "other". The "guilt" stories were scored according to three categories: guilt (if the subject made reference to guilt or if he stated that he was sad/ sorry for the story victim); indifference (if the subject stated that the story character would feel nothing or happiness) and other (any other response). All responses were scored by two judges. Agreement was 100%.

7.3: Results

The data are analyzed separately for each emotion attribution. Table 7.1 displays the results for the happiness and sadness stories. As can be seen the responses of both subject groups were very similar. The dominant emotion attribution to the happiness stories was "happy", the dominant emotion attribution to the sadness stories was "sad". Assuming that emotion attribution ability reflects emotional experience, it can be concluded that APD and CC subjects experience happiness and sadness similarly.

Table 7.1: Emotion attributions for the happiness and sadness stories (as a percentage).

Story type	Group	Emotion attribution	
		Happy	Sad
Happiness	APD	8	0
	CC	8	0
Sadness	APD	0	7
	CC	0	6

Table 7.2 displays the results for the embarrassment stories under the three types of audience condition (no audience, passive audience and negative audience). The pattern of responding was as expected given the findings of Bennett (1989); embarrassment was only attributed if an audience was present. Again the pattern of responding of the two groups was very similar. Assuming that emotion attribution ability reflects emotional experience, it can be concluded that

APD and CC subjects experience embarrassment similarly.

Table 7.2: Emotion attributions for the embarrassment stories under the three types of audience condition (as a percentage).

Audience condition	Group	Emotion attribution	
		Embarrassment	Anger
No audience	APD	0	8
	CC	0	9
Passive audience	APD	10	0
	CC	9	0
Negative audience	APD	7	3
	CC	8	2

Table 7.3 displays the results for the "guilt" stories under the three act conditions (victim + intention; victim - intention; conventional). The emotion attributions of the subjects were scored according to three categories: guilt (responses of guilt, sorry or upset for the victim); indifference (responses of happy or nothing) and other (any other response). As can be seen, the patterns of responding of the two groups do differ, at least when considering the person-harm stories. Fischer's Exact tests comparing APD to CC performance[2] on both of the person-harm stories reveal a significant difference if the harm was intentional (Fischer's Exact; $p < 0.05$) and a trend in the same direction (Fischer's exact; $p < 0.1$) if the harm was not intentional. Figures 7.1 and 7.2 show the attributions of APD and CC subjects for both of these person harm conditions. It can also be seen from these figures (and table 7.3) that these differences were in line with predictions; APD subjects were less likely to show guilt and more likely to show indifference than CC subjects. In addition, it is interesting to note in table 7.3 the lone APD subject who gave guilt attributions to all transgressions; conventional as well as moral. While his presence is not statistically significant it is interesting that this behaviour can be clearly classed (given the definition of guilt described above) as anomalous. It is also interesting to note that, for both groups, intentional violence resulted in significantly less guilt attributions than unintentional violence (McNemar $X^2 = 4.15$, $df = 1$, $p < 0.05$). This is apparently contradictory to the definition of guilt formulated by Hoffman (1987); see section 7.2.3.

[2] Only the differing propensities of the APD and CC subjects to make guilt or indifference attributions was tested. The "other" responses to the intentional harm story (one of embarrassment, the APD, and two of worry, the CCs) were ignored. The reason for this neglect was the ambiguous nature of the subjects' attributed concern; whether it was self or victim directed. It was predicted that an individual with VIM would show victim directed concern while an individual without would show self directed concern.

Table 7.3: Emotion attributions for the "guilt" stories under the three act conditions (scores expressed as a percentage of total scores). Conventional 1 involves talking in class, conventional 2 involves leaving the classroom during a lesson.

	Harm type	Group	Emotion attribution		
			Guilt	Indifference	Other
Intention for violence	Person	APD	1	8	1
		CC	5	3	3
	Object	APD	5	5	0
		CC	5	3	2
No intention for violence	Person	APD	6	4	0
		CC	9	1	0
	Object	APD	5	5	0
		CC	5	5	0
Con	1	APD	1	6	3
		CC	0	6	4
	2	APD	1	7	2
		CC	0	7	3

Figure 7.1: Emotion attributions of APD and CC subjects on the intentional person-harm story.

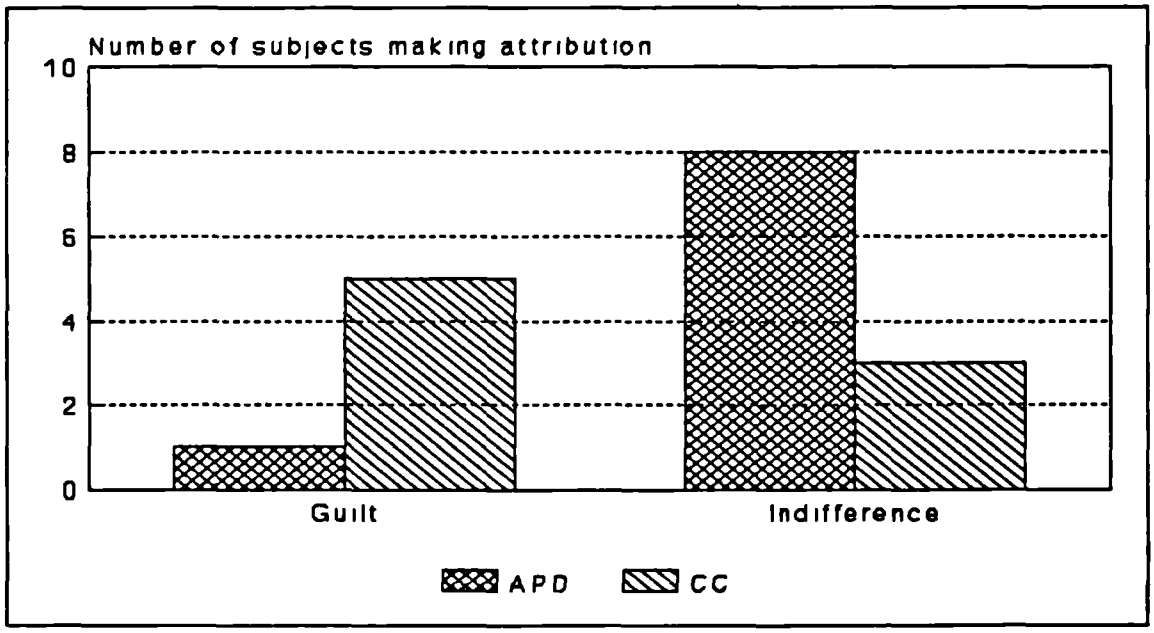
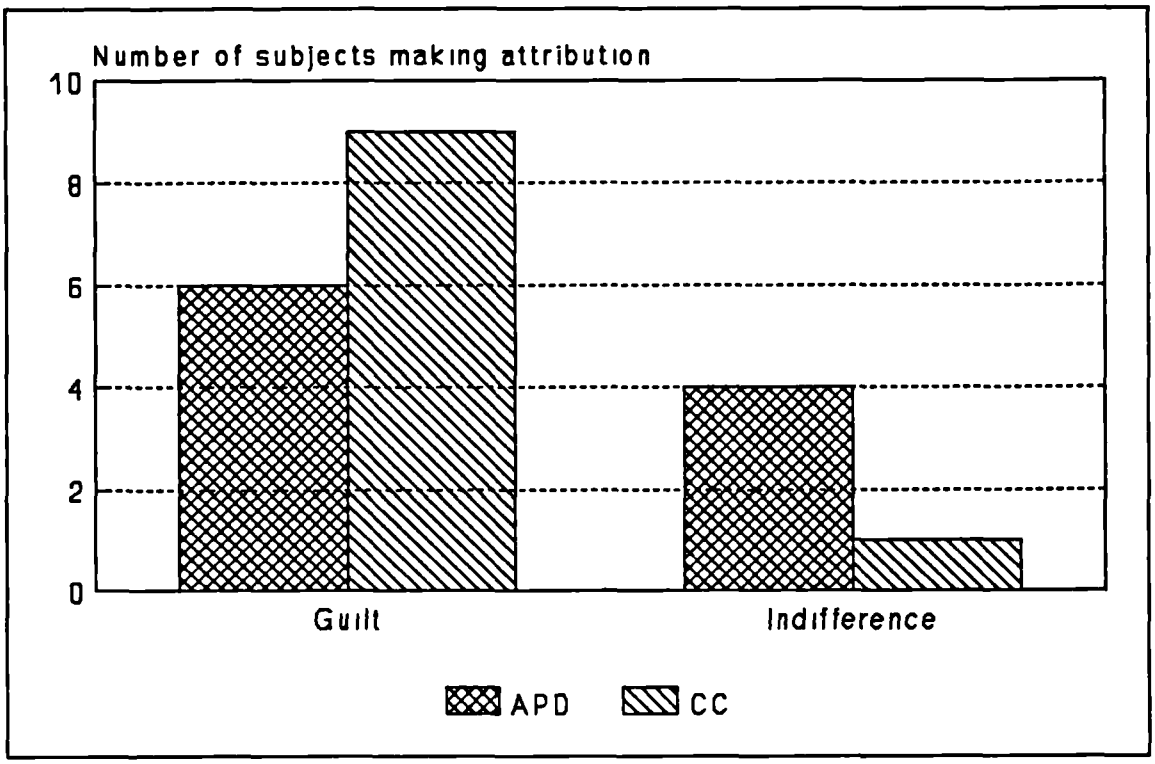


Figure 7.2: Emotion attributions of APD and CC subjects on the unintentional person-harm story.



7.4: Discussion

The present study examined the emotion attributions made by Antisocial Personality disorder (APD) and Criminal Control (CC) subjects. This study revealed: Firstly, that the emotion attributions of APD and CC subjects are very similar if the conditions elicit happiness, sadness or embarrassment attributions. Secondly, that the emotion attributions of APD and CC subjects differ if the conditions involve harm to others. Thirdly, this difference is only present if the victims are present (i.e., there and physically hurt), not inferred (i.e., when they discover their property has been damaged).

The fact that APD and CC subjects attribute happiness, sadness and embarrassment similarly (at least, in the contexts provided here), and assuming emotion attribution ability reflects emotion experience, allows several conclusions to be drawn. The suggestion depicted in the causal model (figure 4.3), that the absence of a prerequisite for the development of all emotions (factor "X") caused APD can be dismissed. If an individual lacked this factor, he would be unable to experience any emotion. However, the fact that the attributions of the APD subjects were only

anomalous for the "guilt" situations, not the "embarrassment", "happiness" or "sadness" attributions indicates that this explanation cannot account for APD.

The fact that APD and CC subjects attribute embarrassment similarly provides additional evidence against a second, rival explanation to the VIM account of APD. As stated in chapter five, Gough (1948) claimed that APD subjects lacked the ability to role take. He claimed that, given this disability, APD subjects would not experience embarrassment. Again, assuming emotion attribution ability reflects emotion experience, and the fact that APD's embarrassment attributions are not anomalous, it can again be concluded (in addition to the evidence provided in chapter five) that APD subjects do not lack the ability to role take.

APD and CC subjects do show differences in their attributions in guilt conditions; APD subjects are less likely to attribute guilt and more likely to attribute indifference. However, these differences only occur if direct victims (people obviously hurt) are present and they are only statistically significant if the harm was intentionally caused. These findings must be explained. First, however, the ability of the APD subjects to refer to guilt must be considered. According to the position explored in this thesis, those without VIM should not possess moral emotions. How then do the APDs, proposed here to be lacking VIM, attribute moral emotions? Unfortunately, there are no adequate accounts of the development of the individual's attribution ability but it seems reasonable to assume that the individual learns to attribute emotion states to others from noting his own emotional reactions to particular environmental situations. But personal experience cannot be the only means available to the individual for gaining information about the attributions of others. Individuals are able to attribute different emotions to different people in the same context. This ability would not be possible if the individual could only use his own experience of emotions induced by that context. Presumably, therefore individuals are learning attribution rules from other sources; i.e., stories told by peers or read or seen on television. APD subjects must have these other sources of information available to them. Thus, individual APD subjects may be able to attribute guilt even though they have never experienced the emotion.

It is probable that the difference between APD and CC responding is only present on the person harm conditions because it is only in these stories that a victim is apparent. It is only in these conditions that a representation of the victim is accessed and an attribution of guilt made. However, an attribution of guilt following victim representations will only be made by

individuals who possess VIM (i.e., the CCs); only they will associate victims (as opposed to just bad behaviour) with guilt. Guilt attributions to the object harm stories by both groups may just be a function of the obviousness of the bad behaviour[3].

The responding of the APDs and the CCs was only significantly different for the intentional harm condition and both subject groups were significantly more likely to make a guilt attribution in the non-intentional harm condition. This effect of intentionality on the responses of the APDs and CCs was not a direct prediction of the VIM position. The VIM position would predict that the individual's arousal level would be determined by the intensity of the distress cues, not by his intention towards the victim. However, this task is not measuring the subject's arousal but his attributions. It is not surprising that the subjects were less likely to attribute guilt in the intentional harm condition. In this condition the subject can make the competing attribution of happiness/ contentment at the success of the plan to hurt the other person. In the unintentional harm condition, there is no competing attribution. The subjects know that hitting someone is a transgression (they stated that this was the case in the moral/ conventional distinction task) and thus if the subject has learnt, through socialisation, a rule such as "*damage* attribute *guilt*" he should attribute guilt. In addition, if the individual has VIM, he should have associated a representation of victim with the attribution of guilt. Thus, the CCs are very likely to attribute guilt in this condition (they know that it is a transgression and an attribution of guilt may be activated by the representation of the victim). The APDs are likely to attribute guilt in the unintentional harm condition (they know that it is a transgression) and there is no competing attribution.

Thus the finding of a difference between the attributions of APD and CC subjects on person damage stories does provide some evidence, albeit indirect, for the VIM model. Obviously, direct evidence is needed. A way of gaining this direct evidence will be described in chapter eight.

7.5: Conclusions

In this chapter I indirectly investigated the emotional capacities of APD and criminal control subjects. I assumed that emotional experience should be reflected in emotional attribution ability. I predicted, following this assumption, that APD but not criminal controls should show impoverished comprehension of the moral emotions. Some evidence was presented that was in line with this prediction.

---

[3] Please note the implicit suggestion that it is not the activation of VIM that is mediating performance on this task. The operation of VIM may play a role, developmentally, in associating victims with guilt but it is not directly responsible for performance on this task. This is in contrast to its role in determining the moral/ conventional distinction. I suggest that it is the aversive arousal generated by VIM as a consequence of representations of moral transgressions which is responsible for the distinction. Whereas object harm stories become, through classical conditioning, conditioned stimuli for the activation of VIM they are not necessarily associated with representations of victims and thus do not necessarily activate attributions of guilt.

Chapter Eight
Conclusions and Future Directions

8.1: Introduction

It is my intention in this last chapter to review the conclusions that can be drawn from the empirical and theoretical work contained in this thesis. These conclusions concern the development of morality (in particular, the development of the moral/ conventional distinction) and the origins of Antisocial Personality Disorder (APD). Following my description of what has been achieved in this thesis, I will then go on to detail some potential future directions for my research. This future work will further explore the hypothesized relationship between the proposed Violence Inhibition Mechanism (VIM) and the development of morality. In addition, this work will investigate the causal model I presented in chapter four that suggested that APD might be a developmental consequence of the lack of VIM.

8.2: Conclusions

This thesis investigated the development of morality; in particular, this thesis addressed the development of the moral/ conventional distinction. As stated in chapter one, the development of this distinction cannot be explained by any of the early formulations of the development of morality; i.e., Kohlberg's (e.g., 1969), Piaget's (1932) or Youniss's (e.g, 1980) accounts. However, models of the distinction are available; those of Turiel and his colleagues (e.g., Smetana, 1983; Turiel, 1977; 1983) and Shweder (e.g., Shweder *et al.*, 1987). At the framework level, both Turiel and his colleagues, and Shweder, advocate that social experiences and self-constructivist processes are prerequisites for the development of the moral/conventional distinction. At the model level, Turiel and his colleagues stress the importance of the different social consequences of moral and conventional transgressions while Shweder stresses the role of care-givers in the construction of normative reality. Only Turiel (1977; 1983) actually specifies some of the self-constructivist processes that might be involved in the construction of the moral/ conventional distinction; role taking, personal experience and counter factual reasoning. Shweder (1990) and Smetana (1983) make reference to them without specifying what they are.

It should be clear that Turiel does not claim that any of the self-constructivist processes or the different social experience with moral and conventional transgressions are *prerequisites* for the

development of the moral/ conventional distinction. However, it seems reasonable to suggest that the mechanisms he lists (role-taking, counter-factual reasoning) must be prerequisites for his account to work. Without the capacity to role take, the observer, from Turiel's position, cannot know of the other's plight[1]. In addition, without the capacity for counter factual reasoning, the observer cannot calculate that it would be "better" if the victim had not been hurt (but see footnote 1).

Of course, as is now clear, to claim that role-taking is a prerequisite for the development of the moral/ conventional distinction is incorrect. Role-taking necessitates the representation of the mental states of others. Autistics are unable to represent the mental states of others (*see* Leslie, 1987; chapter three, this thesis). Autistics are, therefore, unable to role-take. Yet, as was shown in chapter three, despite their inability to role-take, autistic children do distinguish in their judgements between moral and conventional transgressions.

A second problem with Turiel's account and, indeed, with all of the previous models of the development of morality (except Youniss's and Shweder's[2]) is the fact that, despite their claims, they have not accounted for the development of moral judgements. They claim that the child makes moral judgements without explaining why the child should make these judgements. Turiel claims that the child will judge a situation where another is hurt as superior to one where another is not hurt. Both Piaget and Kohlberg claim that representations formed through role-taking are important for the development of concern over the circumstances of others. However, none of these authors consider why an individual should be concerned about the plight of another. It seems that while these previous authors have used a Violence Inhibition Mechanism (VIM) in their own calculations of morality they have not included a VIM as a mechanism in their accounts of moral development.

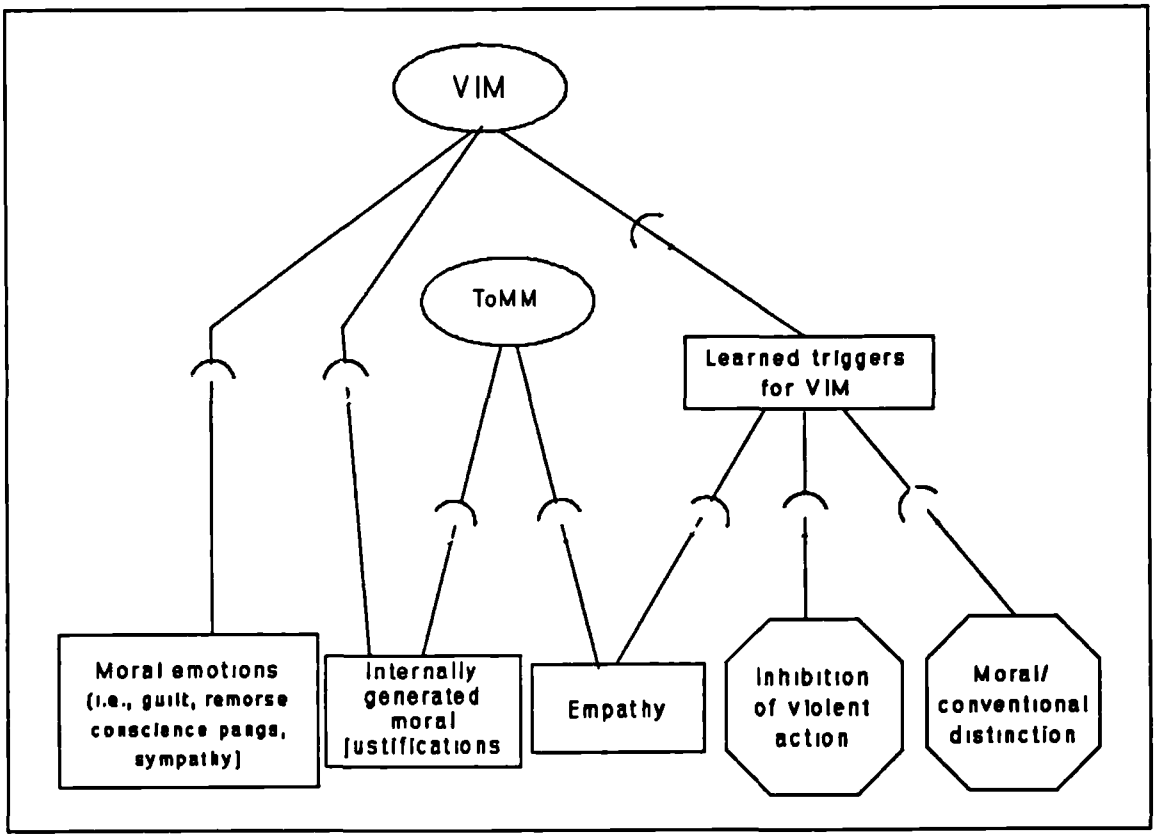
---

[1]Note the fact that Turiel, like Gough (1948; see chapter five) assumes that representations of the others plight, formed by role taking, will, on their own, make moral transgressions seem undesirable. This is despite the fact that there is nothing intrinsic about harm to others that should cause aversion unless a mechanism, such as the proposed Violence Inhibition Mechanism, exists.

[2] Within the models of Shweder and Youniss, there is no need to add a psychological mechanism responsible to explain moral judgements since, according to these authors, judgements are determined by socialisation. They can argue that a child will judge a situation where another is hurt as less desireable than one where the other is not hurt because social pressures have instructed the child to make this judgement. Though, of course, these two models, to be complete, do need a satisfactory explanation of why social pressures should value the plight of the individual.

As I have suggested (see chapters two and four), a Violence Inhibition Mechanism (VIM) is a prerequisite for the development of all moral representations. VIM is conceptualised here as the functional equivalent of the aggression inhibition mechanism that has been hypothesized (e.g., Eibl-Eidesfeldt, 1970; Lorenz, 1966; 1981) to be present in many social animals. These authors suggest that there are mechanisms which result in the termination of an aggressive attack by a conspecific following the production of submission cues by the "victim" animal. VIM is seen as an innately specified schema within contingency scheduling that, when activated by distress cue representations, inhibits other schemata. This inhibition, through inter-schema competition (see chapter two, section 2.4.2, for details of this process), results in the generation of arousal. This arousal, through meaning analysis, may be interpreted as one of the moral emotions (guilt, remorse, sympathy).

Figure 8.1: A DCM of VIM showing the role of ToMM in the development of morality.



In chapter two (figure 2.12), I developed a Developmental Contingency Model (DCM) which suggested that VIM was a prerequisite for the development of the moral emotions, the inhibition of violent action, empathy and the moral/ conventional distinction. In chapter three, this DCM

was expanded (see figure 3.4 represented here as figure 8.1) to show the role of a Theory of Mind Mechanism (ToMM) in the development of morality. ToMM was conceptualised as a prerequisite for empathy (empathy is a consequence of representations formed through role-taking) and internally generated moral justifications. The causal model (see chapter four; figures 4.1 and 4.2) implicit within both these DCMs was tested twice in this thesis. This causal model predicted the cognitive and behavioral consequences of a developmental lack of VIM. Specifically, this model predicted that a developmental lack of VIM would result in: an absence of the moral emotions; a lack of internally generated moral justifications; a failure to inhibit violence and; a failure to make the moral/ conventional distinction. I suggested (chapter four) that Antisocial Personality Disorder (APD) might reflect a developmental lack of VIM. If APD patients do lack VIM, it must be predicted that they will fail to make the moral/ conventional distinction, lack the moral emotions and fail to inhibit violent behaviour. By definition APD patients fail to inhibit their aggressive responses. Chapter four demonstrated that APD patients lack the moral/ conventional distinction and chapter seven provided some evidence that APD subjects lack moral emotions (but not other emotions). This thesis therefore provides evidence that is in line with the DCM developed in chapters two and three and the causal model presented in chapter four. However, this thesis has not provided proof of either model or claim (though see the "future directions" section for details of how to gain such proof).

This thesis also allows two further conclusions. APD is not due either to an inability to role take or to damage to the Supervisory Attentional System (SAS). Chapter five demonstrated that APD patients do not differ from criminal controls in their ability to role take and chapter six demonstrated that APD subjects are not deficient, relative to controls, on tasks assumed to involve the SAS. Indeed, the findings of chapters three, four and five in combination with the evidence of Baron-Cohen *et al.* (1985) that autistics lack a "Theory of Mind" reveals a double dissociation between ability to role take and ability to make the moral/ conventional distinction. Autistics cannot role take but can make the distinction. APD patients can role take but cannot make the distinction.

Thus, in summary, it is clear that role taking is not a prerequisite for the moral/ conventional distinction (though it may be crucial for developing justification categories about the distinction; see chapter three and figure 8.1). It is also clear that VIM may be a prerequisite for the moral/ conventional distinction. It is certainly clear that current formulations need to assume such a mechanism in order to work. There is thus evidence that is compatible with the notion that APD

is a developmental disorder resulting from an absence of VIM though definitive proof has not yet been provided. However, definitive proof has been provided showing that APD is **not** due to an inability to role take or damage to the SAS (as measured by executive function tasks).

8.3: Future Directions

The main area for future research should, of course, be in determining the existence and functioning of VIM. The evidence presented in this thesis is in concordance with the position that VIM is a prerequisite for the development of the inhibition of violent behaviour, the moral emotions and the moral/ conventional distinction. However, there is still no actual direct evidence that VIM exists. This lack must be remedied.

8.3.1: A first study: Determining the operation of VIM

Investigators have reported that subjects show aversive physiological reactions when witnessing the distress of others (Bandura & Rosenthal, 1966; Berger, 1962; Craig & Lowery, 1969; Krebs, 1975). It is assumed here that these subjects are experiencing an arousal response because of the inhibitory activity of VIM (see chapter two). However, the arousal response need not be due to the activation of VIM. This arousal response of observers to others in distress has been used as evidence for empathy (e.g., Batson *et al.*, 1987). It is suggested that observers become aroused from representations of the others condition formed through role-taking. Alternatively, and more in line with the model of emotion presented in chapter two, it might be speculated that these subjects have been predisposed, through cultural pressures, to help others in distress. The activation of a schema to accomplish this goal would, assuming the model of emotion presented in chapter two, result in arousal. A "helping others" schema would be activated which could not be deactivated because, due to the circumstances of the experiment, the subjects were powerless to help. This schema would inhibit all other schemata that might be activated. This inhibition would activate the arousal schema; the arousal would not be due to the activity of VIM.

Thus, while the observation that individuals are aversively aroused by witnessing others in distress is in line with the VIM position, it is not clear evidence that VIM exists. A first study would thus determine the existence of VIM. This study would determine whether there are precise distress cues (e.g., facial expressions and exclamations) that stimulate arousal. Distress cues could be presented to the subject singularly (e.g., a photograph of a sad expression or the

sound of crying) and arousal measured physiologically (e.g., as a galvanic skin response [GSR]).

If there is an arousal response to single distress cue stimuli this would, of course, identify at least some of the specific cues that activate the arousal response; it would determine some of the stimuli that activate VIM. It would also imply, though not prove, that VIM is activated by externally supplied representations and not by representations formed through role-taking. It would be difficult to suggest that the subjects were representing, through role-taking, the plight of an individual when the stimulus material was just a picture of the individual displaying distress cues. In addition, it seems difficult, though not impossible, to argue that a picture of an individual displaying distress cues would excite a "helping schema".

8.3.2: A second study: Determining whether a lack of VIM causes APD

While the first study identified the operating characteristics of VIM, the second study would determine whether APD subjects lack VIM. As stated previously, APD subjects are reported as being without guilt and remorse (DSM-III-R, 1987; Hare, 1980). The DCM presented in chapters two and three (see figure 3.4) suggests that one of the causes of the lack of the development of guilt is the lack of VIM. A second study would test this position.

The first study will have identified two sets of stimulus cues; one set that induces arousal in the normal subject (distress cues), one set that has no arousal consequences. A second study would use these cues to test the hypothesis that the antisocial personality lacks the mechanism which generates the aversive reaction to distress cues. It is predicted that the antisocial personality will not be aroused by distress cues.

Obviously, the discovery of cues that elicit an arousal response allows a superior form of diagnosis of APD. APD would be diagnosed on the basis of a cognitive deficit (a lack of VIM) rather than the collection of behaviours that Hare's (1980; 1985) Psychopathy Checklist (PCL) or DSM-III-R supply. Behavioral clusters may be due to more than one bio-cognitive cause. Diagnosing on the basis of cognitive deficit allows the selection of a group which is homogeneous at least according to one factor.

Further studies might determine whether other clinical populations possess VIM. In particular, a study might determine whether autistic subjects show arousal to distress cues. Following the

DCM developed in chapter three (figure 3.4), it would be predicted that autistic subjects would show arousal to distress cues. There is no reason to believe that autistics lack VIM. Indeed, given that autistic subjects make the moral/ conventional distinction, there is every reason to believe that autistic subjects possess VIM. It would thus be predicted that autistic subjects would show an arousal response to distress cues. If they do not show arousal to distress cues, the operation of VIM cannot be responsible for the moral/ conventional distinction.

The finding of arousal to distress cues in autistics would also allow conclusions about the operation of VIM. As stated above, previous researchers have taken the evidence of arousal response to the distress of others as evidence for empathy. However, autistic subjects must be incapable of empathy since autistic subjects are unable to role-take. Thus, demonstrating an arousal response to distress cues by autistic subjects will clearly demonstrate that the operation of VIM is not mediated by representations formed through role-taking.

8.3.3: The development of VIM

At present VIM is conceptualised as a cognitive mechanism which inhibits behaviour in the presence of distress cues. This inhibition, through the inter-schema competition process (see chapter two), indirectly results in arousal. Additional studies might determine the development of VIM.

It is unclear at what age VIM becomes on-line. It would be expected that VIM is functioning from an early age and, indeed, some evidence indicates that this is the case; infants cry to the sound of crying. Simner (1971) reported this reactive crying in 2- to 3- day old children, a finding that has been replicated by Sagi & Hoffman (1976). This reactive crying is not simply a response to noxious auditory stimuli; infants do not cry to equally loud and intense non-human sounds (Simner, 1971). Of course, this reactive crying may not be mediated by VIM[3], however, it remains that it would be expected that VIM is on-line from an early age; the moral conventional distinction is present in individuals from the age of 36 months. By looking at the VIM functioning of children at different ages it would be possible to determine at what age VIM came on-line.

[3]This proposition can be tested, assuming APD patients prove to lack VIM, by examining whether APD subjects show arousal to the sound of a babies crying. If they do, the sound of a babies crying cannot be mediated by VIM.

A second question concerns the development of VIM. In the DCM, represented in figure 8.1, I claimed that VIM is a prerequisite for the development of learned triggers for VIM. In chapter two, I suggested that, possibly through a process of classical conditioning, conceptual "victim" representations might become triggers for VIM activation in addition to distress cues. However, I have not yet determined that there are learned triggers which activate VIM - it is necessary to show arousal as a consequence of a "victim" representation; i.e., it is necessary to demonstrate arousal in the listener to moral, but not conventional, transgression stories. If individuals do show arousal to moral transgression stories, it would then be possible to determine the age at which this arousal response appears.

8.2.4: VIM and moral development

It would be interesting to examine the relationship between VIM, APD and subject performance on the moral/ conventional distinction. It would be expected that a lack of VIM would correlate well with APD diagnosis on the basis of the Hare's (1980; 1985) Psychopathy Checklist (PCL). It would also be expected that a lack of VIM would be highly associated with an absence of the moral/ conventional distinction. Indeed, according to the DCM presented in chapter two (see figure 8.1), it is predicted that those without VIM should not make the moral/ conventional distinction. A future experiment might therefore investigate compare the moral/ conventional distinction of those who show an arousal response to distress cues and those who do not. It would be interesting to examine the moral/ conventional distinctions of those subject's who were classed as APD by Hare's PCL but who possessed VIM according to the results of the arousal experiment and also those subjects who were classed as criminal control according to Hare's PCL but who did not possess VIM according to the arousal experiment. It would be predicted that the possession of VIM should be a better prediction of the presence of the moral/ conventional distinction than Hare's PCL score.

An additional study might investigate whether VIM is involved in the encoding of moral transgressions. As stated in chapter one, Arsenio and Ford (1985; study 2) observed that inducing a negative emotional state increased the recall of moral transgressions but did not effect the recall of conventional transgressions. I suggest that the operation of VIM generates the negative affect that is encoded in the heading of the record containing the representation of the transgression. VIM would only be activated by moral transgressions. It would thus be predicted, if APD patients are shown to lack VIM, that inducing negative affect in APD patients

would not increase recall of moral transgressions.

8.4: A concluding comment

In this thesis, I have attempted to explore morality by investigating the amoral. I have developed an account which describes morality as a consequence of a basic inhibitory mechanism. Morality is seen here neither as a direct product of "rational thought" nor of emotion but as a developmental consequence of this innately specified violence inhibition mechanism (VIM). VIM is the prerequisite for moral representation and moral emotion.

References

American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders (1st edn) (DSM-I).* APA Washington, D.C.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders (2nd edn) (DSM-II).* APA Washington, D.C.

American Psychological Association. (1987). *Diagnostic and statistical manual of mental disorders (3rd revised edition) (DSM-III-R).* APA, Washington, DC,

Amnesty International. (1973). *Report on torture.* Duckworth, London.

Appellof, E.S. & Augustine, E.A. (1985). Prefrontal functions in juvenile delinquents. *Journal of clinical and experimental neuropsychology*, **7**, 79-109.

Arsenio, W.F. & Ford, M.E. (1985). The Role of Affective Information in Social-Cognitive Development: Children's Differentiation of Moral and Conventional Events. *Merrill-Palmer Quarterly*, **31**, 1-17.

Bandura, A. & Rosenthal, T.L. (1966). Viacarous classical conditioning as a function of arousal level. *Journal of Personality and Social Psychology*, **3**, 54-62.

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of autism and developmental disorders*, **18**, 379-402.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, **30**, 285-297.

Baron-Cohen, S., Leslie, A.M. & Frith, U. (1985). Does the autistic have a "Theory of Mind"? *Cognition*, **21**, 37-46.

Baron-Cohen, S., Leslie, A.M. & Frith, U. (1986). Mechanical, behavioral and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, **4**, 113-125.

Batson, C.D. (1988). Prosocial Motivation: Is it ever truly altruistic? In *Advances in Experimental Social Psychology*, (Berkowitz, L., ed.), Academic Press, New York.

Batson, C.D., Fultz, J. & Schoenrade, P.A. (1987). Adults' emotional reactions to the distress of others. In *Empathy and its development* (Eisenberg, N. and Strayer, J., eds.), Cambridge University Press, Cambridge.

Bennett, M. (1989). Children's self-attribution of embarrassment. *British Journal of Developmental Psychology* **7**, 207-217.

Bennett, M. & Gillingham, K. (1990). The role of self-focused attention in children's attributions of social emotions to the self. *Journal of Genetic Psychology*, **7**, 207-219.

Berger, S.M. (1962). Conditioning through vicarious instigation. *Psychological Review*, **69**, 450-466.

Berscheid, E. (1982). Attraction and emotion in interpersonal relationships. In *Affect and cognition: The Seventeenth Annual Carnegie Symposium on Cognition* (Clark, M.S. and Fiske, S.T., eds.), Erlbaum, Hillside, N.J.

Berscheid, E. (1983). Emotion. In *Close relationships* (Kelley, H.H., Berscheid, E., Christensen, A., et al, eds.), Freeman, San Francisco.

Bisiach, E., Capitani, E., Luzzatti, C. & Perani, D. (1981). Brain and conscious representation of outside reality. *Neuropsychologia*, **19**, 543-551.

Blackburn, R. (1988a). On morla judgements and personality disorders: The myth of psychopathic personality revisited. *British Journal of Psychiatry*, **153**, 505-512.

Blackburn, R. (1988b). Psychopathy and personality disorder. In *Adult abnormal psychology* (Miller, E. and Cooper, P.J., eds.), Churchill Livingstone, Edinburgh.

Blasi, A. (1980). Bridging moral cognition and moral action: a critical review of the literature. *Psychological bulletin*, **88**, 1-45.

Bowler, D.M. (1989). Theory of mind in Asperger's syndrome. Paper presented at the London

conference of the British Psychological Society.

Bryson, S.E., Clark, B.S. & Smith, I.M. (1988). First report of a Canadian epidemiological study of autistic syndromes. *Journal of Child Psychology and Psychiatry*, **29**, 433-445.

Carter, D.B. & Patterson, C.J. (1982). Sex roles as social conventions: The development of children's conceptions of sex-role concepts. *Developmental Psychology*, **18**, 812-824.

Cleckley, H. (1964). *The mask of sanity*. Mosby, New York.

Colby, A. & Kohlberg, L. (1987). *The measurement of moral judgement*. Cambridge University Press, New York,

Craig, K.D. & Lowrey, J.H. (1969). Heart rate components of conditioned vicarious autonomic responses. *Journal of Personality and Social Psychology*, **11**, 381-387.

Damon, W. (1977). *The social world of the child*. Jossey-Bass, San Francisco.

Davidson, P., Turiel, E. & Black, A. (1983). The effect of stimulus familiarity on the use and criteria of justifications in children's social reasoning. *British Journal of Developmental Psychology*, **1**, 49-65.

Dawson, G. & Fernald, M. (1987). Perspective-taking ability and its relationship to the social behavior of autistic children. *Journal of autism and developmental disorders*, **17**, 487-498.

DeGelder, B. (1987). On not having a theory of mind. *Cognition*, **27**, 285-290.

Dewey, J. (1895). The theory of emotion: II. The significance of emotions. *Psychological Review*, **2**, 13-32.

Duncan, J. & Humphreys, G.W. (1989). Visual search and stimulus familiarity. *Psychological Review*, **96**, 433-458.

Dunn, J. & Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology*, **23**, 791-798.

Dworkin, R. (1978). *Taking rights seriously*. Harvard University Press, Cambridge, Mass.

Edwards, C.P. (1987). Culture and the construction of moral values: A comparative ethnography of moral encounters in two cultural settings. In *The emergence of morality in young children* (Kagan, J. and Lamb, S., eds.), The University of Chicago Press, Chicago.

Eibl-Eibesfeldt, I. (1970). *Ethology: The biology of behaviour*. Holt, Rinehart and Winston, New York.

Eisenmajer, R. & Prior, M. (1991). Cognitive linguistic correlates of "Theory of Mind" ability in autistic children. *British Journal of Developmental Psychology*, **9**, 351-364.

Ekman, P., Friesen, W.V. & Ellsworth, P. (1972). What emotion categories or dimensions can observers judge from facial behaviour. In *Emotion in the human face* (Ekman, P., ed.), Cambridge University Press, Cambridge.

Ekman, P., Levenson, R.W. & Friesen, W.V. (1983). Autonomic nervous activity distinguishes among emotions. *Science*, **221**, 1208-1210.

Elliot, F.A. (1978). Neurological aspects of antisocial behavior. In *The psychopath* (Reid, W.H., ed.), Bruner/ Mazel, New York.

Eysenck, H.J. (1964). *Crime and personality*. Routledge & Kegan Paul, London.

Fabes, R.A., Eisenberg, N., McCormick, S.E. & Wilson, M.S. (1988). Preschoolers' attributions of the situational determinants of naturally occuring emotions. *Developmental Psychology*, **24**, 376-385.

Feshbach, N.D. (1978). Studies of empathic behavior in children. In *Progress in experimental personality research* (Maher, B.A., ed.), Academic Press, New York.

Feshbach, N.D. (1983). Learning to care: A positive approach to child training and discipline. *Journal of Clinical Child Psychology*, **12**, 266-271.

Fodor, E.M. (1973). Moral development and parent behaviour antecedents in adolescent psychopaths. *Journal of genetic psychology*, **122**, 37-43.

Frankenhaeuser, M. (1971). Experimental approaches to the study of human behaviour as related to neuroendocrine functions. In *Society, stress and disease* (Levi, L., ed.), Oxford University Press, London.

Frankenhaeuser, M. (1975). Experimental approaches to the study of catecholamines and emotion. In *Emotions: Their parameters and measurement* (Levi, L., ed.), Raven Press, New York.

Fridlund, A. (1991a). Darwin's anti-Darwinism in the expression of the emotions in man and animals. *International Review of Emotion*, **2**.

Fridlund, A. (1991b). The sociality of solitary smiling: Potentiation by an implicit audience. Unpublished manuscript.

Frijda, N.H. (1986). *The emotions*. Cambridge University Press, Cambridge.

Frith, U. (1989). *Autism: Explaining the enigma*. Blackwell, Oxford.

Frith, U., & Happé., F.G.E. (1992). Knowing that you think and feeling that you sense in autism. Unpublished manuscript.

Frith, U., Morton, J. & Leslie, A. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neuroscience*, **14**, 433-438.

Fry, P.S. & Ogsten, D.G. (1971). Emotion as a function of the labeling of interuption produced arousal. *Psychonomic science*, **24**, 153-154.

Gabennesch, H. (1990). The perception of social conventionality by children and adults. *Child Development*, **61**, 2047-2059.

Gewirth, A. (1978). *Reason and morality*. University of Chicago Press, Chicago.

Gibbs, J.C. (1987). Social processes in delinquency: The need to facilitate empathy as well as sociomoral reasoning. In *Moral development through social interaction* (Kurtines, W.M. and

Gewirtz, J.L., eds.), Wiley, New York.

Gibson, J. & Haritos-Fatouros, M. (1986). The education of a torturer. *Psychology today*, **20**, 50-58.

Gnepp, J., McKee, E. & Domanic, J.A. (1987). Children's use of situational information to infer emotion: Understanding emotionally equivocal situations. *Developmental Psychology*, **23**, 114-123.

Gorenstein, E.E. (1982). Frontal lobe functions in psychopaths. *Journal of Abnormal Psychology*, **91**, 368-379.

Gorenstein, E.E. & Newman, J.P (1982). Disinhibitory Psychopathology: A new perspective and a model for research. *Psychological Review*, **37**, 301-315.

Gough, H.G. (1948). A sociological theory of psychopathy. *American Journal of sociology*, **53**, 359-366.

Gough, H.G. (1969). *Manual for the California Psychological Inventory*. Consulting Psychologists Press Palo Alto, CA.

Happé, F.G.E. (1991). *Theory of mind and communication in autism.* Unpublished Ph.D. thesis.

Hare, R.D. (1970). *Psychopathy: Theory and research.* Wiley New York.

Hare, R.D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and individual differences*. **1**, 111-119.

Hare, R.D. (1983). Diagnosis of antisocial personality disorder in two prison populations. *American Journal of Psychiatry*, **140**, 887-890.

Hare, R.D. (1984). Performance of psychopaths in cognitive tasks related to frontal lobe function. *Journal of Abnormal Psychology*, **93**, 133-140.

Hare, R.D. (1985a). A comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology*. **53**, 7-16.

Hare, R.D. (1985b). Scoring manual for the Psychopathy Checklist. Unpublished manuscript.

Hare, R.D. & Cox, D.N. (1978). Clinical and empirical conceptions of psychopathy, and the selection of subjects for research. In *Psychopathic behaviour: Approaches to research* (Hare, R.D. and Schalling, D., eds.), Wiley, Chichester, England.

Haritos-Faroutos, M. (1983). *Antecedent conditions leading to the bahaviour of a torturer: Fallacy or reality?* Unpublished Ph.D. thesis. University of Thessaloniki, Greece.

Haritos-Faroutos, M. (1988). The official torturer: A learning model for obedience to authority of violence. *Journal of applied social psychology*, **18**, 1107-1120.

Hart, S.D., Forth, A.E. & Hare, R.D. (1990). Performance of criminal psychopaths on selected neuropsychological tests. *Journal of Abnormal Psychology*, **99**, 374-379.

Helwig, C.C., Tisak, M.S. & Turiel, E. (1990). Children's social reasoning in context: Reply to Gabennesch. *Child Development*, **61**, 2068-2070.

Hoffman, J.J., Hall, R.W. & Bartsch, T.W. (1987). On the relative importance of "psychopathic" personality and alchoholism on neuropsyhcological measures of frontal lobe destruction. *Journal of Abnormal Psychology*, **96**, 158-160.

Hoffman, M.L. (1970). Conscience, personality and socialiszation techniques. *Human Development*, **13**, 90-126.

Hoffman, M.L. (1975). Developmental synthesis of affect and cognition and its implications for altruistic motivation. *Developmental Psychology*, **11**, 607-622.

Hoffman, M.L. (1977). Empathy, its development and prosocial implications. In *Nebraska Symposium on Motivation (Vol. 25)* (Keasey, C.B., ed.), University of Nebraska Press, Lincoln, pp. 169-218.

Hoffman, M.L. (1984). Empathy, its limitations, and its role in a comprehensive moral theory. In *Morality, moral development, and moral behavior* (Gewirtz, J. and Kurtines, W., eds.), Wiley, New York, pp. 283-302.

Hoffman, M.L. (1987). The contribution of empathy to justice and moral judgment. In *Empathy and its development* (Eisenberg, N. and Strayer, J., eds.), Cambridge University Press,

Cambridge.

Hollos, M., Leis, P. & Turiel, E. (1986). Social reasoning children and adolescents. *Journal of Cross Cultural Psychology*, **17**, 352-374.

Home office/ Department of health and social security. (1975). *Report of the committee on abnormal offenders.* HMSO, London.

Houghton, G., & Tipper, S.P. (1992a). A model of inhibitory mechanisms in selective attention. In *Inhibitory mechanisms in attention, memory and language.* (Dagenbach, D. and Carr, T., eds.), Academic Press, London.

Houghton, G., & Tipper, S.P. (1992b). Inhibitory control fo neurodynamics: Opponent mechanisms in sequencing and selective attention. In *Neurodynamics and psychology.* (Oaksford, M. & Brown, G.D.A., eds.), Academic Press, London.

Hughes, C.H. & Russell, J. (1990) One reason why deception is difficult for autistic subjects: Failure to disengage from an object. Unpublished manuscript.

Izard, C.E. (1971). *The face of emotion.* Appleton-Century-Crofts, New York.

Izard, C.E. (1977). *Human emotions.* Plenum Press, New York.

Johnson-Laird, P.N. (1983a). *Mental models: Towards a cognitive science of language, inference and consciousness.* Cambridge University Press, Cambridge.

Johnson-Laird, P.N. (1983b). A computational analysis of consciousness. *Cognition and Brain Theory*, **6**, 499-508.

Jurkovic, G.J. & Prentice, A. (1977). Relation of moral and cognitive development to dimensions of juvenile delinquency. *Journal of Abnormal Psychology* **86**, 414-420.

Kagan, J. (1984). *The nature of the child.* Basic Books New York,

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, **2**, 217-250.

Kanner, L. & Eisenberg, L. (1956). Early infantile autism 1943-1955. *American Journal of Orthopsychiatry*, **26**, 55-65.

Karpman, B. (1941). On the need for seperating psychopathy into two distinct types: The symptomatic and the Idiopathic. *Journal of Criminal Psychopathy*, **3**, 112-137.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In *Handbook of socialisation theory and research* (Goslin, D.A., ed.), Rand McNally, Chicago.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In *Cognitive development and epistemology* (Mischel, T., ed.), Academic Press, New York.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In *Moral development and behaviour: Theory, research, and social issues* (Lickona, T., ed.), Holt, Rinehart & Winston, New York.

Kohlberg, L. (1981). The philosophy of moral development: Moral stages and the idea of justice. In *Vol 1. of Essays on moral development* (Harper and Row, San Francisco.

Kohlberg, L. (1984). *Essays on moral development. Vol I: The philosophy of moral development.* Harper & Row, San Francisco.

Kohlberg, L. & Kramer, R. (1969). Continuities and disconuities in childhood and adult moral development. *Human Development*, **12**, 93-120.

Kohlberg, L., Levine, C. & Hewer, A. (1983). *Moral stages: A current formulation and a response to critics.* Karger Basil.

Komatsu, L.K. & Galotti, K.M. (1986). Children's reasoning about social, physical, and logical regularities: A look at two worlds. *Child Development*, **57**, 413-420.

Krebs, D.L. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, **32**, 1134-1146.

Laupa, M. & Turiel, E. (1986). Children's conceptions of adult and peer authority. *Child Development*, **57**, 405-412.

Lee, M. & Prentice, N.M. (1988). Interrelations of empathy, cognition, and moral reasoning with dimensions of juvenile delinquency. *Journal of Abnormal Child Psychology*, **16**, 127-139.

Leslie, A.M. (1987). Pretense and representation: The origins of "Theory of Mind". *Psychological Review*, **94**, 412-426.

Leslie, A.M. (1988). Some implications of pretense for mechanisms underlying the child's theory of mind. In *Developing Theories of the Mind* (Astington, P.L., Harris, P.L. and Olson, D.R., eds.), Cambridge University Press, Cambridge.

Leslie, A.M. & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, **6**, 315-324.

Lezak, M. (1976). *Neuropsychological assessment.* Oxford University Press, New York.

Lifton, R.J. (1986). *The nazi doctors: Medical killing and the psychology of genocide.* Basic books, New York.

Lorenz, K. (1966). *On aggression.* Harcourt Brace Jovanovich, New York.

Lorenz, K. (1981). *The foundation of ethology.* Springer-Verlag, New York.

Lotter, V. (1966). Epidemiology of autistic conditions in young children: I. Prevalence. *Social Psychiatry*, **1**, 124-137.

Lotter, V. (1974). Factors related to outcome in autistic children. *Journal of autism and Childhood Schizophrenia*, **4**, 263-277.

Lunney, G.H. (1970). Using analysis of variance with a dichotomous dependent variable: an empirical study. *Journal of Educational Measurement*, **7**, 263.

Lykken, D.T. (1957). A study of anxiety in the sociopathic personality. *Journal of abnormal and social psychology*, **55**, 6-10.

Mandler, G. (1964). The interuption of behaviour. In *Nebraska Symposium on Motivation* (Levine, D., ed.), University of Nebraska Press, Lincoln, Neb.

Mandler, G. (1975). *Mind and emotion.* Wiley, New York.

Mandler, G. (1984). *Mind and Body.* Norton, New York.

Mandler, G. (1987). A constructivist theory of emotion. In *Psychological and biological approaches to emotion* (Stein, N.S., Leventhal, B.L. and Trabasso, T., eds.), Hillsdale, NJ.

Mandler, G. (1991). Emotions, evolution, and aggression: Myths and conjectures. In *International review of studies on emotion. Vol. II* (Strongman, K.T., ed.), John Wiley & Sons, Chichester.

Manstead, A.S.R. & Wagner, H.L. (1981). Arousal, cognition and emotion: An appraisal of two-factor theory. *Current psychological reviews*, **1**, 35-54.

Marshall, J.G. & Halligan, P.W. (1988). Blaindsight and insight in visuo-spatial neglect. *Nature*, **336**, 766-769.

McCleary, R.A. (1966). Response-modulating function of the limbic system: Initiation and suppression. In *Progress in physiological psychology.* (Stellar, E. & Sprague, J.M. eds.), Academic Press, New York.

Meyer, L.B. (1956). *Emotion and meaning in music.* University of Chicago Press, Chicago.

Milgram, S. (1974). *Obedience to authority.* Harper and Row, New York.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology*, **23**, 53-57.

Millon, T. (1983). *Millon clinical multiaxial inventory manual.* Interpretive scoring systems, Minneapolis.

Milner, B. (1964). Some effects of frontal lobectomy in man. In *The frontal granular cortex and behaviour* (Warren, J.M. and Akert, K., eds.), McGraw-Hill, New York.

Moffitt, T.E. & Henry, B. (1989). Neuropsychological assessment of executive functions in self-reported delinquents. *Development and psychopathology*, **1**, 105-118.

Morton, J. (1986). Developmental Contingency Modelling. In *Theory building in developmental psychology* (van Geert, P.L.C., ed.), Elsevier, Amsterdam.

Morton, J. & Frith, U. (*in press*). Causal modelling: A structural approach to developmental psychopathology. In *Manual of developmental psychopathology*, Vol I. (Cicchetti, D., & Cohen, D.H. eds.). Wiley, New York.

Morton, J., Hammersley, R.H. & Bekerian, D.A. (1985). Headed records: A model for memory and its failure. *Cognition*, **20**, 1-23.

Much, N. & Shweder, R. (1978). Speaking of rules: The analysis of culture in breach. In *New directions for child development. Vol. 2, Moral Development* (damon, W., ed.), Jossey-Bass, San Francisco.

Murdock, G.P. (1980). *Theories of illness: A world survey*. University of Pittsburgh Press, Pittsburgh.

Newman, J.P., & Kosson, D.S. (1986). Passive avoidance learning in psychopathic and non-psychopathic offenders. *Journal of abnormal psychology*, **95**, 257-263.

Newman, J.P., Patterson, C.M., & Kosson, D.S. (1987). Response perseveration in psychopaths. *Journal of abnormal psychology*, **96**, 145-148.

Newman, J.P., Widom, C.S., & Nathan, S. (1985). Passive avoidance in syndromes of disinhibition: Psychopathology and extraversion. *Journal of personality and social psychology*, **48**, 1316-1327.

Nisan, M. (1987). Moral norms and social conventions: A cross-cultural comparison. *Child Development*, **23**, 719-725.

Norman, D. & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In *Consciousness and self-regulation* (Davidson, R.J., Schwartz, G.E. and Shapiro, D., eds.), Plenum Press, New York.

Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, **52**, 114-121.

Nucci, L.P. (1982). Conceptual development in the moral and conventional domains: Implications for values eductation. *Review of educational research* **52**, 93-122.

Nucci, L.P. & Nucci, M. (1982a). Children's social interactions in the context of moral and conventional transgressions. *Child Development* **53**, 403-412.

Nucci, L.P. & Nucci, M. (1982b). Children's responses to moral and conventional transgressions in free-play settings. *Child Development* **52**, 1337-1342.

Nucci, L.P. & Turiel, E. (1978). Social interactions and the development of social concepts in preschollo children. *Child Development* **49**, 400-407.

Nucci, L., Turiel, E. & Encarnacion-Gawrych, G.E. (1983). Social interactions and social concepts: Analysis of morality and convention in the Virgin Islands. *Journal of Cross Cultural Psychology* **14**, 469-487.

Nunez, M. & Riviere, A. (1990). Theory of mind and other cognitive developments. Paper presented at IV European Conference of Developmental Psychology, at Stirling, Scotland.

Oatley, K. (1988). Plans and the communicative function of emotions: A cognitive theory. In *Cognitive perspectives on emotion and motivation* (Hamilton, V., ed.), Kluwer Academic Publishers, New York.

Oatley, K. & Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion* **1**, 29-50.

Ozonoff, S., Rogers, S.J. & Pennington, B.F. (1992). Asperger's syndrome: Evidence for an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry*.

Partridge, G.E. (1930). Current conceptions of psychopathic personality. *American Journal of Psychiatry*. **10**, 53-99.

Paulhan, F. (1930). *The laws of feeling*. Kegan Paul, Trench, Trubcr London.

Perner, J., Frith, U., Leslie, A.M. & Leekam, S.R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Development* **60**, 689-700.

Perner, J. & Wimmer, H. (1985). John thinks that Mary thinks that ... " Attribution of second-order beliefs by 5-10 year old children. *Journal of Experimental Child Psychology* **39**, 437-471.

Perry, D.G. & Perry, L.C. (1974). Denial of suffering in the victim as a stimulus to violence in aggressive boys. *Child Development* **45**, 55-62.

Piaget, J. (1932). *The Moral Development of the Child.* Routledge and Kegan Paul, London.

Pichot, P. (1978). Psychopathic behavior: Approaches to research. In *Psychopathic behavior: a Historcal review - reverse* (Hare, R.D. and Schalling, D.S., eds.), Wiley, Chichester.

Pontius, A.A. & Yudowitz, B.S. (1980). Frontal lobe system dysfunction in some criminal actions as shown in the narratives test. *Journal of nervous and mental disease* **168**, 111-117.

Raine, A. (1985). A psychometric assessment of Hare's checklist for psychopathy on an English prison sample. *British Journal of Clinical Psychology* **24**, 247-258.

Rawls, J. (1971). *A theory of justice.* Harvard University Press Cambridge, Mass.

Reed, T. & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of autism and developmental disorders* **20**, 555-568.

Reisenzein, R. (1983). The Schacter theory of emotion: Two decades later. *Psychological bulletin* **94**, 239-264.

Rest, J.R. (1983). Morality. In *Handbook of Child Psychology* (Mussen, P.H., ed.), John Wiley, New York.

Russell, J., Mauthner, N., Sharpe, S. & Tidswell, T. (1991). The "windows" task as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology* **9**, 331-349.

Sagi, A. & Hoffman, M.L. (1976). Empathic distress in newborns. *Developmental Psychology* **12**, 175-176.

Samenov, S.E. (1984). *Inside the criminal mind.* Random House New York,

Schalling, D. (1978). Psychopathy-related personality variables and the psychophysiology of socialization. In *Psychopathic behavior: Appraches to research* (Hare, R.D. and Schalling, D., eds.), Wiley, Chichester, England.

Schmauk, P.J. (1970). Punishment, arousal, and avoidance learning in psychopaths. *Journal of Abnormal Psychology* **76**, 325-335.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B* **298**, 199-209.

Shallice, T. (1988a). *From neuropsychology to mental structure.* Cambridge University Press, Cambridge.

Shallice, T. (1988b). Information-processing models of consciousness: possibilities and problems. In *Consciousness in contemporary science* (Marcel, A.J. and Bisiach, E., eds.), Oxford University Press, Oxford.

Shallice, T. & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, **114**, 727-741.

Shallice, T. & Burgess, P. (*in press*). Higher-order cognitive impairments and frontal-lobe lesions in man. In *frontal lobe functions and injury* (Levin, H., Eisenberg, H.M. and Benton, A.L., eds.), Oxford University Press, Oxford.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex* **14**, 294-303.

Sher, M.A. (1971). Pupillary dilation before and after interuption of familiar and unfamiliar sequences. *Journal of Personality and Social Psychology* **20**, 281-286.

Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing:

II. Perceptual learning, automatic attending, and a general theory. *Psychological Review* **84**, 127-190.

Shweder, R.A., Mahapatra, M. & Miller, J.G. (1987). Culture and moral development. In *The emergence of morality in young children* (Kagan, J. and Lamb, S., eds.), University of Chicago Press, Chicago, pp. 1-83.

Shweder, R.A. & Much, N.C. (1987). *Moral development through social interaction.* Wiley, New York.

Siegal, M. & Storey, R.M. (1985). Day care and childrens's conceptions of moral and social rules. *Child Development* **56**, 1001-1008.

Simner, M.L. (1971). Newborn's response to the cry of another infant. *Developmental Psychology* **5**, 136-150.

Skoff, B.F. & Libon, D.J. (1987). Impaired executive functions in a sample of male juvenile offenders. *Journal of experimental and clinical neuropsychology* **9**, 60.

Smetana, J. (1981). Preschool children's conceptions of moral and social rules. *Child Development* **52**, 1333-1336.

Smetana, J. (1983). Social-cognitive development: Domain distinctions and coordinations. *Developmental Review* **3**, 131-147.

Smetana, J. (1984). Toddlers' social interactions regarding moral and conventional transgressions. *Child Development* **55**, 1767-1776.

Smetana, J. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology* **21**, 18-29.
Smetana, J. (1986). Preschool children's conceptions of sex-role transgressions. *Child Development* **57**, 862-871.

Smetana, J.G. & Braeges, J.L. (1990). The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly* **36**, 329-346.

Smetana, J., Bridgeman, D.L. & Turiel, E. (1983). Differentiation of domains and prosocial behaviour. In *The nature of prosocial development: Interdisciplinary theories and strategies* (Bridgeman, D.L., ed.), Academic Press, New York.

Smetana, J., Kelly, M. & Twentyman, C.T. (1984). Abused, neglected and nonmaltreated children's conceptions of moral and social-conventional transgressions. *Child Development* **55**, 277-287.

Sodian, B. & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry*.

Song, M., Smetana, J.G. & Kim, S.Y. (1987). Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology* **23**, 577-582.

Sperber, D. & Wilson, D. (1981). Irony and the use-mention distinction. In *Radical pragmatics* (Cole, P., ed.), Academic press, New York.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition.* Blackwell Oxford.

Spitzer, R.L., Endicott, J. & Robins, E. (1978). Research diagnostic criteria. *Archives of general psychiatry* **35**, 773-782.

Staub, E. (1990). The psychology and culture of torture and torturers. In *Psychology and torture* (Suedfeld, P., ed.), Hemisphere, New York.

Steffenberg, S. & Gillberg, C. (1986). Autism and autism-like conditions in Swedish rural and urban areas: a population study. *British journal of psychiartry* **149**, 81-87.

Steffenberg, S. & Gillberg, C. (1986). Autism and autism-like conditions in Swedish rural and urban areas: a population study. *British Journal of Psychiatry* **149**, 81-87.

Stoddart, T. & Turiel, E. (1985). Children's concepts of cross-gender activities. *Child Development* **56**, 1241-1252.

Strayer, J. (1986). Children's attributions regarding the situational determinants of emotion in self and others. *Developmental Psychology* **22**, 649-654.

Taylor, D.A. & Harris, P.L. (1983). Knowledge of the link between emotion and memory among normal and maladjusted boys. *Developmental Psychology* **19**, 832-838.

Tisak, .M. & Turiel, E. (1984). Children's conception of moral and prudential rules. *Child Development* **55**, 1030-1039.

Tisak, M.S. & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology* **24**, 352-357.

Trasler, G.B. (1973). Criminal behaviour. In *Handbook of abnormal psychology* (Eysenck, H.J., ed.), Pitman, London.

Trevathan, S. & Walker, L.J. (1989). Hypothetical versus real-life moral reasoning among psychopathic and delinquent youth. *Development and psychopathology* **1**, 91-103.

Turiel, E. (1977). Distinct concpetual and developmental domains: Social convention and morality. In *Nebraska Symposium on Motivation* (Howe, H.E. and Keasey, C.B., eds.), University of Nebraska Press, Lincoln.

Turiel, E. (1978). The development of concepts of social structure. In *The development of social understanding* (Glick, J. and Clarke-Stewart, A., eds.), Gardner, New York.

Turiel, E. (1983). *The development of social knowledge: Morality and Convention.* Cambridge University Press Cambridge.

Turiel, E. (1983). Domains and categories in social cognitive development. In *The relationship between social and cognitive development* (Overton, W., ed.), Erlbaum, Hillside, N.J.

Turiel, E. & Davidson, P. (1986). Heterogeneity, inconsistency, and asynchrony in the development of cognitive structures. In *Stage and Structure: Reopening the Debate* (Levin, I., ed.), Ablex Press, Norwood,N.J.

Turiel, E., Killen, M. & Helwig, C.C. (1987). Morality: Its structure, functions, and vagaries. In *The emergence of morality in young children* (Kagan, J. and Lamb, S., eds.), University of Chicago Press, Chicago, pp. 155-245.

Turiel, E. & Smetana, J. (1984). Social knowledge and action: The coordination of domains. In *Morality, moral behavior and moral development: Basic issues in theories and research* (Kurtines, W.M. and Gewirtz, J.L., eds.), John Wiley, New York.

Vaillant, G.E. (1975). Sociopathy as a human process: a viewpoint. *Archives of general psychiatry* **32**, 178-183.

Weiskrantz, L. (1986). *Blindsight.* Clarendon Press Oxford.

Weston, D. & Turiel, E. (1980). Act-rule relations: Children's concepts of social rules. *Developmental Psychology* **236**, 417-424.

Widom, C.S. (1976). Interpersonal and personal construct systems in psychopaths. *Journal of consulting and clinical psychology* **44**, 614-623.

Williams, L. (1992). Torture and the torturer. *Psychologist* **5**, 305-308.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and the constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**, 103-128.

Winer, B.T. (1971). *Statistical principles in experimental design.* McGraw-Hill New York,

Wing, L. (1981). Language, social and cognitive impairments in autism and severe mental retardation. *Journal of autism and developmental disorders* **11**, 31-44.

Wing, L. (1988). The continuum of autistic charecteristics. In *Diagnosis and assessment in autism* (Schopler, E. and Mesibov, G.B., eds.), Plenum, New York,

Wing, L. & Gould, J. (1978). Systematic recording of behaviors and skills of retarded and psychotic children. *Journal of autism and Childhood Schizophrenia* **8**, 79-97.

Wing, L. & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of autism and developmental disorders* **9**, 11-29.

Wing, L., Gould, J., Yeates, S.R. & Brierley, L.M. (1977). Symbolic play in severely retarded and autistic children. *Journal of Child Psychology and Psychiatry* **18**, 167-178.

Wing, L. & Wing, J.K. (1971). Multiple impairments in early childhood autism. *Journal of autism and Childhood Schizophrenia* **1**, 256-266.

Wong, S. (1985). The criminal and institutional behaviours of psychopaths. Program branch user report, Ministry of the solicitor General of Canada, Ottawa, Canada.

World Health Organization. (1987). *Mental disorders: A glossary and guide to their classification in accordance with the 10th revision of the International Classification of Diseases (ICD-10).* World Health Organization Geneva,

Wulff, S.B. (1985). The symbolic and object play of children with autism: A review. *Journal of autism and developmental disorders* **15**, 139-148.

Yochelson, S. & Samenov, S.E. (1977). *The criminal personality (Vol. 2).* Aronson New York,

Youniss, J. (1980). *Parents and peers in social development.* University of Chicago Press Chicago,

Youniss, J. (1981). Moral development through a theory of social construction. *Merrill-Palmer Quarterly* **27**, 385-403.

Youniss, J. (1987). Social construction and moral development: Update and expansion of an idea. In *Moral development through social interaction* (Kurtines, W.M. and Gewirtz, J.L., eds.), Wiley, New York.